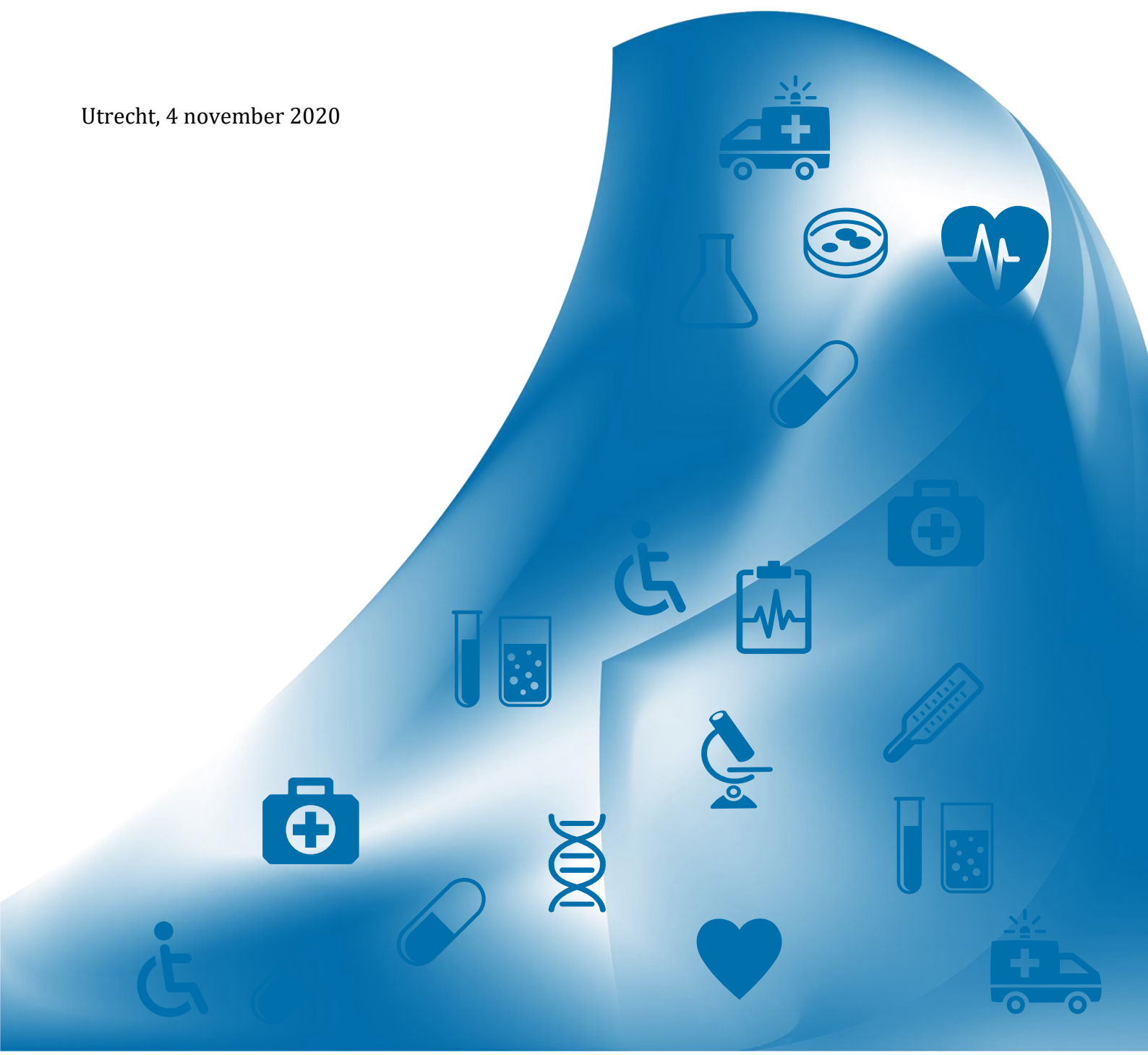


NULMETING DATABESCHIKBAARHEID IN GEZONDHEID EN ZORG

Utrecht, 4 november 2020



SAMENVATTING

Het ministerie van Volksgezondheid, Welzijn en Sport heeft Capgemini Invent gevraagd om te inventariseren welke initiatieven en/of constructies in Nederland bestaan om gezondheidsdata breder beschikbaar te maken voor het ontwikkelen, trainen, valideren en toepassen van artificiële intelligentie. Deze inventarisatie geeft een beeld van alle huidige initiatieven en/of constructies die in en voor Nederland bestaan om gezondheidsdata beschikbaar te maken en om datasolidariteit vorm te geven. Daarnaast geeft deze inventarisatie inzicht in welke richting de huidige initiatieven zich bewegen en met welk einddoel.

Deze inventarisatie richt zich specifiek op dataconstructies die gezondheidsdata vanuit meerdere bronnen ontsluiten en beschikbaar maken, waarbij voor databeschikbaarheid de definitie is gehanteerd dat er voldoende variëteit en kwaliteit aan data beschikbaar, toegankelijk, veilig en praktisch is om kwalitatief goede AI te ontwikkelen en toe te passen. Er zijn 23 initiatieven meegenomen in deze inventarisatie¹. Op basis van verdiepende vragenlijsten en gesprekken is een samenhangend beeld gevormd van het huidige landschap én de belangrijkste vraagstukken omtrent databeschikbaarheid in gezondheid en zorg.

De belangrijkste inzichten vanuit dit onderzoek worden hieronder toegelicht:

- Uit de inventarisatie onder dataconstructies die bijdragen aan de databeschikbaarheid voor gebruik van AI in gezondheid en zorg komt naar voren dat er geen gebrek is aan gezondheidsdata, maar wel aan toegankelijke gezondheidsdata. De focus in het veld ligt op dit moment nog vooral op het verzamelen, verwerken/bewerken en beschikbaar stellen van de data via samenwerkingsconstructies of het ontsluiten van additionele bronnen.
- Voor de partijen in deze inventarisatie is het vergroten van de databeschikbaarheid een belangrijk doel. Het gaat hierbij niet om meer data te verzamelen op één plek, maar om gezondheidsdata toegankelijk en interoperabel te maken. Vanwege de huidige regelgeving die toeziet op zorgvuldig en doelbewust delen van gezondheidsdata, maar ook initiatieven waarmee decentraal opgeslagen data kan worden gebruikt (bijvoorbeeld de Personal Health Train²) is het ook niet nodig om data in toenemende mate centraal te verzamelen.
- De focus van databeschikbaarheid in het veld ligt minder bij de zoektocht naar toegevoegde waarde van AI in de praktijk. AI wordt op dit moment in het veld gezien als de volgende logische stap na brede databeschikbaarheid van gezondheidsdata. De databeschikbaarheid die wordt geboden door deze constructies kan gebruikt worden voor AI, maar vergt in veel gevallen een lange termijnvisie voordat waardevolle AI kan worden ontwikkeld en toegepast met deze data.
- Het handelingsperspectief om de databeschikbaarheid voor gebruik van AI te vergroten ligt binnen de volgende thema's :
 - **Samenbrengen van initiatieven** door middel van het ondersteunen van constructen en (koepel)organisaties die het gefragmenteerde landschap proberen te harmoniseren met gelijke standaarden en procedures.
 - **Verbreden van databeschikbaarheid** door het opzetten van een centraal aanspreekpunt in het veld zodat nieuwe initiatieven gemakkelijker kunnen starten en bestaande initiatieven de complementariteit met andere partijen kunnen vergroten.
 - **Verbinding met de industrie** opzoeken zodat de databeschikbaarheid ook ten goede komt aan partijen die de capaciteiten hebben om artificiële intelligentie te ontwikkelen en naar de praktijk te brengen binnen het domein gezondheid en zorg.
 - **Vertrouwen uit de samenleving** opwekken door burgers en patiënten te informeren over de impact van het delen van gezondheidsgegevens voor de ontwikkeling, training en validatie van innovatieve toepassingen.

¹ In de periode van 4 augustus t/m 24 september (2020) heeft Capgemini Invent gezocht naar data constructies in Nederland en deze constructies geïnterviewd. Ondanks de inspanningen om een zo groot mogelijk deel van alle constructies te vinden, is het mogelijk dat er een aantal ontbreken.

² Personal Health Train (2020), <https://pht.health-ri.nl/nl>

Inhoudsopgave

1.	Inleiding	4
2.	Projectaanpak	5
2.1	Stappenplan.....	5
2.2	Onderzoeksscope.....	6
2.3	Classificatie.....	8
3.	Databeschikbaarheid	9
3.1	Vraagstukken databeschikbaarheid.....	9
3.2	Uitdagingen voor gezondheidsdata.....	10
3.3	Gerelateerde Onderzoeken.....	12
4.	Resultaten	13
4.1	Overzichtsmatrix.....	14
4.2	Kengetallen.....	15
4.3	Beschrijvende informatie.....	16
4.3.1	Doelgroepen.....	16
4.3.2	Zorgsectoren.....	17
4.3.3	Financiering.....	17
4.3.4	Samenwerkingsmodellen.....	17
4.3.5	Specifieke aandoeningen.....	18
4.3.6	Waarde van data voor AI “aan het bed”.....	18
4.4	Datavraagstukken.....	19
4.4.1	Domein van verwerking.....	19
4.4.2	Security & Privacy.....	19
4.4.3	Conformiteit en vertrouwen.....	20
4.4.4	Betrouwbaarheid en kwaliteit.....	20
4.5	Voorwaarden voor databeschikbaarheid.....	21
4.5.1	Regelgeving.....	21
4.5.2	Harmonisatie.....	22
4.5.3	Vertrouwen.....	22
5.	Conclusie & aanbevingen	24
5.1	Samenbrengen van initiatieven.....	24
5.2	Verbreden databeschikbaarheid.....	25
5.3	Verbinding met de industrie.....	26
5.4	Vertrouwen uit de samenleving.....	26
6.	Overzicht initiatieven	28
7.	Appendix	39
7.1	Referenties.....	39

1. INLEIDING

Artificiële Intelligentie (AI) heeft de potentie om een enorme impact op de kwaliteit van onze zorg en gezondheid te hebben. Door de kracht van AI te gebruiken om zelf analyses te maken, uitkomsten te voorspellen en daarvan te leren, kunnen we tot op maat gemaakte adviezen of acties komen. Denk aan snellere diagnoses, preventie, nauwkeurigere behandelingen en beter ingelichte patiënten. Zo kan AI een behandelende arts de mogelijkheden bieden om “aan het bed” een weloverwogen beslissing te maken op basis van vergelijkbare ervaringen uit andere zorginstellingen. Tegenwoordig zijn er al een groot aantal AI-initiatieven, van voorspellingsmodellen voor ziektebeelden tot het bieden van een persoonlijke behandeling op maat³

Hoewel AI veel kansen biedt voor de toekomst, zijn er ook een aantal zaken waar rekening mee gehouden moet worden. Een toekomst voor gebruik van AI binnen de zorg begint bij de beschikbaarheid van kwalitatief goede en volledige data. Waarbij de beschikbaarheid van gezondheidsdata over verschillende organisatie en zorginstellingen heen één van de grootste uitdagingen vormt. Daarom heeft het programma ‘Waardevolle AI voor gezondheid’ van het ministerie van Volksgezondheid, Welzijn en Sport (VWS) Capgemini Invent gevraagd om te inventariseren welke initiatieven en/of constructies in Nederland bestaan om gezondheidsdata breder beschikbaar te maken voor het gebruik van Artificiële Intelligentie (AI). Het gaat hierbij specifiek om dataconstructies die medisch relevante data vanuit meerdere bronnen ontsluiten en beschikbaar maken. Naast het inzicht in welke initiatieven er zijn, moet de inventarisatie duidelijkheid bieden over mogelijke (in de toekomst) noodzakelijke aanvullingen om gebruik van AI in de zorg te stimuleren.

Dit rapport bouwt voort op actuele informatie en rapporten over databeschikbaarheid en voegt extra waarde toe door een overkoepelend beeld te schetsen van het huidige landschap van databeschikbaarheid in gezondheid en zorg. Dit onderzoek beschrijft welke partijen in Nederland bijdragen aan het beschikbaar stellen van gezondheidsdata ten behoeve van AI⁴, de richting waarop het veld zich ontwikkelt en de overkoepelende thema’s waar deze partijen tegen aanlopen. Zo wordt een samenhangend beeld gevormd van het huidige landschap, de toekomstvisie én de belangrijkste vraagstukken omtrent databeschikbaarheid voor AI in de gezondheidszorg in Nederland.

Leeswijzer

De nulmeting databeschikbaarheid is opgebouwd uit onderstaande hoofdstukken:

Hoofdstuk 2: *De projectaanpak, de projectscope en de classificatie per constructie*

Hoofdstuk 3: *Vraagstukken rondom databeschikbaarheid die zijn meegenomen in deze inventarisatie en zijn besproken met de verschillende constructies*

Hoofdstuk 4: *De resultaten van deze inventarisatie uitgesplitst op hoofdthema’s*

Hoofdstuk 5: *De conclusies en bijbehorende discussie op basis van de resultaten*

Hoofdstuk 6: *Een korte samenvatting per dataconstruct uit de inventarisatie*

Hoofdstuk 7: *De appendix met hierin de referenties naar ondersteunende documenten*

³ KPMG (2020), “Inventarisatie AI-toepassingen in gezondheid en zorg in Nederland”

⁴ In de periode van 4 augustus t/m 24 september (2020) heeft Capgemini Invent gezocht naar data constructies in Nederland en deze constructies geïnterviewd. Ondanks de inspanningen om een zo groot mogelijk deel van alle constructies te spreken, is het onvermijdelijk dat er mogelijk een aantal ontbreken.

2. PROJECTAANPAK

Het doel van deze nulmeting is om initiatieven op het gebied van databeschikbaarheid voor gezondheid en zorg in kaart te brengen voor ontwikkeling van AI, zowel vanuit de industrie, wetenschap, instituten, samenwerkingsverbanden als mogelijke andere invalshoeken. De nulmeting moet inzicht geven in de huidige initiatieven in het veld, welke richting deze initiatieven zich bewegen en met welk einddoel. De nulmeting is een eerste stap in het komen tot een voorstel hoe in Nederland databeschikbaarheid voor AI op een verantwoorde wijze geborgd kan worden.

2.1 STAPPENPLAN

Om de hierboven benoemde doelen van deze nulmeting te bereiken is de inventarisatie uitgevoerd middels de volgende vier stappen:



Figuur 1 - Stappenplan nulmeting databeschikbaarheid in gezondheid en zorg

Stap 1: Verkennen

In de eerste stap is onderzoek gedaan naar welke initiatieven zich momenteel bezighouden met het vergroten van de databeschikbaarheid voor de zorg in Nederland. Voor de inventarisatie van initiatieven is gebruik gemaakt van een aantal verschillende bronnen:

- 1. Input van het ministerie van Volksgezondheid, Welzijn en Sport (VWS):**
Gezien de rol en ervaring van het ministerie van VWS vormde een lijst met initiatieven aangeleverd door het ministerie het startpunt van het onderzoek
- 2. Netwerk Capgemini Invent en deskresearch:**
Aanvulling van deze groslijst met initiatieven geïdentificeerd middels deskresearch en via contacten en ervaring binnen Capgemini Invent
- 3. Suggesties vanuit initiatieven zelf:**
Gebruik maken van de collectieve kennis in het veld gedurende de verdiepende gesprekken
- 4. Online webformulier:**
Opzetten van een online webformulier die is gedeeld via social media, de website datavoorgezondheid.nl en de Capgemini Invent website, waarbij initiatieven zichzelf konden aanmelden voor dit onderzoek.

Uit deze vier bronnen is een lijst met data initiatieven in Nederland gekomen van 63 initiatieven. Hierbij moet worden benoemd dat dit geen uitputtende lijst is. Er is geprobeerd alle relevante initiatieven in kaart te brengen, maar het is onvermijdelijk dat enkele constructies niet zijn meegenomen in deze inventarisatie. Hiernaast is in kaart gebracht welke databeschikbaarheidsvraagstukken relevant zijn in het algemeen en specifiek binnen het gezondheid en zorgdomein (zie hoofdstuk 3).

Stap 2: Verzamelen

Alle in stap 1 verzamelde initiatieven zijn beoordeeld op basis van de scopingcriteria (hoofdstuk 2.2) om te bepalen of deze initiatieven, projecten of organisaties relevant zijn voor deze inventarisatie. Op basis hiervan is een voor dit onderzoek relevante lijst van constructen opgesteld, die vervolgens gecontacteerd zijn voor een verdiepend interview.

Stap 3: Verdiepen

Tijdens deze stap zijn alle constructies benaderd voor verdiepende interviews⁵. Om efficiënt informatie te verzamelen over alle constructies is een digitale vragenlijst opgesteld en gedeeld met contactpersonen vanuit de dataconstructie voorafgaand aan het interview. Deze vragenlijst werd voor het interview ingevuld teruggestuurd, zodat vragen gericht gesteld konden worden.

Kenmerken per construct zijn verzameld en gerelateerd aan algemene informatie over de constructie, de achterliggende datatechniek, security & privacy, volwassenheid, schaalbaarheid en toekomstbeeld. Tijdens deze interviews is inzichtelijk gemaakt in welke richting de huidige initiatieven zich bewegen en met welk einddoel. Daarbovenop was het van belang om inzichtelijk te hebben welke initiatieven complementair zijn en mogelijk (in de toekomst) elkaar kunnen aanvullen.

Niet alle initiatieven die binnen de scope van dit onderzoek vallen zijn meegenomen in deze inventarisatie, omdat niet alle partijen zijn ingegaan op het verzoek om deel te nemen (zie sectie 2.2).

Stap 4: Verwerken

Aan de hand van de bevindingen van de interviews zijn de datavraagstukken, obstakels van schaalbaarheid en toekomstvisie van de constructen in kaart gebracht. De in stap 3 verzamelde classificatiepunten maken het makkelijker om in deze stap de verwante constructies met elkaar te vergelijken en verbanden te leggen door de constructies en initiatieven te groeperen. Om beter inzicht te krijgen in het datalandschap dient namelijk ook over de constructies heen gekeken te worden. Er is een inschatting gemaakt van de huidige vraagstukken met betrekking tot databeschikbaarheid en kwaliteit en in hoeverre deze met de bestaande constructies worden opgelost. Daaruit is een eerste aanzet voortgekomen voor mogelijke richtingen voor borging van het gebruik van AI in gezondheid en zorg.

2.2 ONDERZOEKSSCOPE

Het in kaart brengen van de initiatieven die bijdragen aan de beschikbaarheid van data voor secundaire toepassingen is de focus van dit project. Om te beoordelen of een initiatief bijdraagt aan data beschikbaarheid, door data uit verschillende bronnen beschikbaar te maken voor AI zijn vier scopingcriteria vastgesteld:



Bijdrage aan de databeschikbaarheid

Alle huidige initiatieven en constructies in ontwikkeling die zorgen dat voldoende variëteit en kwaliteit aan data beschikbaar, toegankelijk, veilig en praktisch is om kwalitatief goede AI te ontwikkelen en te gebruiken door data uit verschillende bronnen beschikbaar te maken



Geografische relevantie

Alle constructies in Nederland of in internationale samenwerking met Nederland



Medische relevantie

Initiatieven die zijn opgezet of worden toegepast in de Nederlandse cure, care en preventie



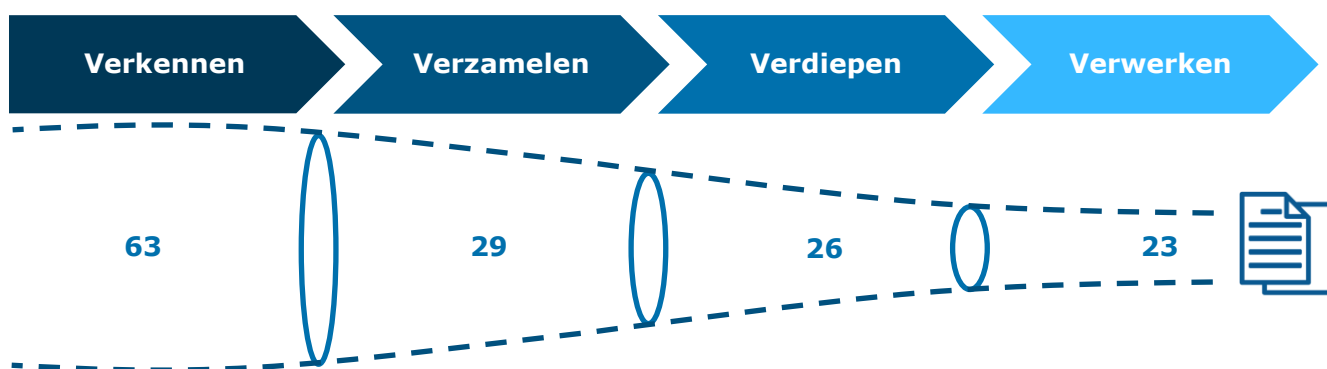
Solidariteit buiten beheergebied

Constructies die de data beschikbaar stellen buiten het eigen beheergebied

⁵ Op basis van de beschikbare informatie in de eerste fase van het onderzoek zijn enkele constructies ten onrechte geselecteerd voor verdere verdieping. Tijdens het contact met deze dataconstructie via de vragenlijst of een interview zijn deze constructies alsnog buiten de scope van dit onderzoek en de inventarisatie gehouden.

Elk initiatief is beoordeeld op deze scopingcriteria en zodoende wel of niet meegenomen in het onderzoek. De lijst van 63 constructies die uit stap 1 zijn voortgekomen, werd na in acht nemen van de scopingcriteria gereduceerd tot een lijst van 29 initiatieven. De voornaamste redenen waardoor initiatieven buiten scope vielen was omdat ze behoorde tot één van onderstaande voorbeelden:

1. Ondersteunende partijen die technologische oplossingen bieden om data te analyseren of inzichtelijk te maken en niet zelf data beschikbaar stellen.
2. Partijen die voor specifiek wetenschappelijk onderzoek data verzamelen en alleen ter hergebruik aan wetenschappelijk onderzoek beschikbaar stellen
3. Registraties van metadata of administratieve informatie uit het zorg domein.
4. Ondersteunende partijen die zich bezighouden met best practices voor dataconstructies
5. Geografisch irrelevante partijen



Figuur 2 - Aantal constructies per stap in de inventarisatie

Van de overige 29 initiatieven zijn 23 meegenomen in deze inventarisatie. De overige constructies zijn niet meegenomen in deze inventarisatie vanwege geheimhoudingsplicht, geen interesse of geen response op ons verzoek om deel te nemen.

2.3 CLASSIFICATIE

Om de resultaten van de inventarisatie van in-scope initiatieven overzichtelijk te presenteren is een classificatie essentieel. Door de constructies in te delen op hun eigenschappen zijn zij makkelijk vindbaar, vergelijkbaar en de resultaten waardevoller doordat verbanden gelegd kunnen worden door de constructies en initiatieven te groeperen. Om beter inzicht te krijgen in het landschap van gezondheidsdata initiatieven is ook over de constructies heen gekeken. Welke vraagstukken worden (potentieel) opgelost door de bestaande constructies? Hoe zijn de constructies complementair aan elkaar? Wat is het einddoel van deze constructies? In welke richting liggen mogelijke oplossingen voor nog bestaande uitdagingen?

De definitieve classificatie lijst is verdeeld in de volgende categorieën:

1. **Details van de beheerder:** *het gaat hier om bijvoorbeeld contactgegevens, type bedrijf en locatie.*
2. **Details van de constructie/initiatief:** *de basisgegevens van de constructie worden hier genoemd (bijbehorende webpagina, de status en deelnemende partijen etc.) alsmede welke vraagstukken de constructie helpt oplossen, andere betrokken partijen en een beschrijving van de impact die de constructie heeft.*
3. **Details partij bron data:** *het gaat hier om gegevens over de partij die de data beschikbaar stelt, bijvoorbeeld het type partij en de manier waarop data geleverd wordt aan het construct.*
4. **Beschikbaarstelling data (bronnen & techniek):** *in deze categorie gaat het om de data infrastructuur en de volwassenheid daarvan. Deze technische aspecten vormen uitdagingen voor het delen van data, dus hoe constructies die aanpakken levert waardevolle lessen.*
5. **Details bron data:** *details bron data creëert inzicht in de data zelf, bijvoorbeeld de grote van de dataset, is de data gericht op een specifiek ziektebeeld en wat voor soort data is het (tekstueel, numeriek, afbeeldingen etc.)*
6. **Data – Waarde voor AI:** *om inzichtelijk te maken welke waarde de constructies toevoegen aan de ontwikkeling van AI, worden in deze categorie vragen gesteld om de datakwaliteit te beoordelen.*
7. **Security & privacy:** *inzicht te vergaren in hoe de databeveiliging achter de constructie in elkaar zit. Hoe wordt de data beschermd en welke voorwaarden zitten er bijvoorbeeld aan de toegang? Daarnaast gaat het hier om toestemming van patiënten en ethische vraagstukken.*
8. **Schaalbaarheid en toekomstbestendigheid:** *naast de technische kwaliteiten, zijn ook de randvoorwaarden essentieel voor het ontwikkelen van een volwassen, schaalbare en toekomstbestendige constructie. Er zijn vragen gesteld of er bijvoorbeeld barrières bestaan voor het opschalen van de constructie, wat de belangrijkste voorwaarden om complementariteit met andere initiatieven te vergroten zijn en wat de toekomstplannen zijn.*

3. DATABESCHIKBAARHEID

Met behulp van AI, die is ontwikkeld en getraind op kwalitatief goede en integrale data, kan meer waarde voor patiënten, zorgverleners en burgers uit gezondheidsdata gehaald worden. Denk bijvoorbeeld aan modellen die het herstel na operatie voorspellen en preventieve interventies zoals fysiotherapie adviseren of een algoritme die op basis van de radiologiebeelden longpatiënten in een vroeg stadium kan diagnosticeren op basis van de longscans van andere patiënten. Hiervoor is het wel van belang dat de onderliggende data integraal beschikbaar is, en van goede kwaliteit. Dit begint bij de patiënten en burgers die er vertrouwen in moeten hebben dat er zorgvuldig met hun data wordt omgegaan. Dit vertrouwen onder patiënten is aanwezig zo blijkt uit een onderzoek van de Patiënten Federatie Nederland naar patiënten waar lichaamsmateriaal is afgenomen. Van de onderzoeksgroep kijkt 83% positief aan tegen het gebruik van medische data voor wetenschappelijk onderzoek. Echter laat dit onderzoek ook zien dat slechts aan 26% van de patiënten de vraag gesteld is om informatie te delen. Een gemiste kans, want zo wordt een groot aandeel gezondheidsdata niet vastgelegd en kan deze data niet gebruikt worden voor secundaire toepassingen die van toegevoegde waarde kunnen zijn.

In de praktijk gaat het bij het gebruik van medische gegevens veelal over geanonimiseerde of gepseudonimiseerde gezondheidsdata. Voorbeelden hiervan zijn diagnose- en behandelingsinformatie, zorgdeclaraties of informatie over de effectiviteit van een behandeling. Veel van deze gezondheidsdata wordt nu wel vastgelegd, maar blijft binnen de muren van een specifieke organisatie. Hierdoor wordt het gebruik van AI voor gezondheid complexer, omdat je veel data nodig hebt van verschillende instellingen met andere type patiënten om betrouwbare en objectieve algoritmes te ontwikkelen. Zogenaamde multi-center data biedt hier een oplossing doordat de data vanuit verschillende organisaties wordt ontsloten waardoor er meer testcases zijn waarop een algoritme kan leren. Deze multi-center data is een belangrijke voorwaarde voor de ontwikkeling van AI om ervoor te zorgen dat de voorspellende waarde van het algoritme zo objectief en betrouwbaar mogelijk is. Zo zal bijvoorbeeld een algoritme die uitsluitend getraind is op data van mannelijke patiënten uit landelijke gebieden naar verwachting een minder goed voorspellend vermogen hebben als deze wordt toegepast op vrouwelijke patiënten uit de stad. Dit onderschrijft het belang om vanuit meerdere instellingen gegevens te verzamelen van diverse ziektebeelden om op deze manier bij te dragen aan de databeschikbaarheid ten behoeve van de ontwikkeling van kwalitatief goede algoritmes. Met deze inventarisatie moet duidelijk worden welke data initiatieven en/of constructies er zijn die hieraan bijdragen en data vanuit meerdere bronnen ontsluiten dan wel beschikbaar maken. Denk hierbij aan samengestelde data van instellingen of vanuit samenwerkingsverbanden.

In dit hoofdstuk is eerst ingegaan op de belangrijkste databeschikbaarheidsvraagstukken in het algemeen (sectie 3.1) en de specifieke beschikbaarheidsvraagstukken binnen het zorgdomein (sectie 3.2). Dit is relevant omdat deze vraagstukken zijn meegenomen in de verdieping van de constructie tijdens de inventarisatie. De thema's uit deze secties keren terug in de resultaten en conclusie van dit rapport. Dit hoofdstuk sluit af met een beknopte samenvatting van andere onderzoeken die in het verlengde liggen van deze inventarisatie (sectie 3.3).

3.1 VRAAGSTUKKEN DATABESCHIKBAARHEID

Databeschikbaarheid is van essentieel belang om goede kwalitatieve AI te ontwikkelen. Hierbij gaat het niet alleen om de vindbaarheid en beschikbaarheid van data, maar ook over thema's als kwaliteit, betrouwbaarheid, volledigheid, volume en de relevantie van de verzamelde data voor verschillende AI toepassingen. Dataconstructies dragen bij aan bovenstaande kwesties door een volledig beeld te schetsen van alle beschikbare data gerelateerd aan een bepaald onderwerp of thema. Ten grondslag van databeschikbaarheid liggen nog enkele vraagstukken die belangrijk zijn om te benoemen.

Volgens de Nederlandse AI Coalitie zijn een gebrek aan **datasoevereiniteit**⁶, **vertrouwen** en **veiligheid** over het gebruik van gegevens momenteel de grootste belemmering voor het grootschalig delen van gegevens⁷. Deze drie bouwstenen zijn meegenomen tijdens het onderzoek, om een beter beeld te schetsen van de huidige situatie omtrent het beschikbaar stellen van data voor AI.

Om datasoevereiniteit te beschermen, en vertrouwen en veiligheid te creëren zijn er voorwaarden en juridische- en ethische standaarden aan datahergebruik gesteld rondom datadeling en gegevenskoppeling. Deze standaarden zijn bedoeld om datadeling te vergroten door datasoevereiniteit, vertrouwen en veiligheid beter te waarborgen. Echter blijft sommige data van dusdanige waarde dat datadeling niet in het belang is van personen of organisaties. Voorbeelden hiervan zijn data die privacy gevoelig zijn of waar een concurrentievoordeel mee te behalen valt. Daarnaast kan een verschillende interpretatie van standaarden ertoe leiden dat data uit voorzorg niet wordt gedeeld om fouten te voorkomen. Informatie-uitwisseling en koppeling over domeinen heen wordt hierdoor lastiger in het huidige versnipperde landschap, waarin nog geen systeem van landelijke datadeling is⁸.

3.2 UITDAGINGEN VOOR GEZONDHEIDSDATA

In de sectie hierboven zijn datasoevereiniteit, vertrouwen en veiligheid genoemd als drie generieke uitdagingen voor databeschikbaarheid. Als het gaat om het gebruik van gezondheidsdata voor AI-toepassingen zijn er potentiële datavraagstukken die hier aan verwant zijn of waar volgens dit onderzoek meer nadruk op moet komen te liggen. Deze uitdagingen zijn dan ook meegenomen tijdens deze inventarisatie en getoetst tijdens de verdieping met de verschillende dataconstructies. Wij hebben de volgende vier uitdagingen geïdentificeerd aan de start van dit onderzoek:



Domein van verwerking



Security & Privacy



Conformiteit en vertrouwen



Betrouwbaarheid en kwaliteit

Figuur 3: Parameters meegenomen in het onderzoek

1. Domein van verwerking

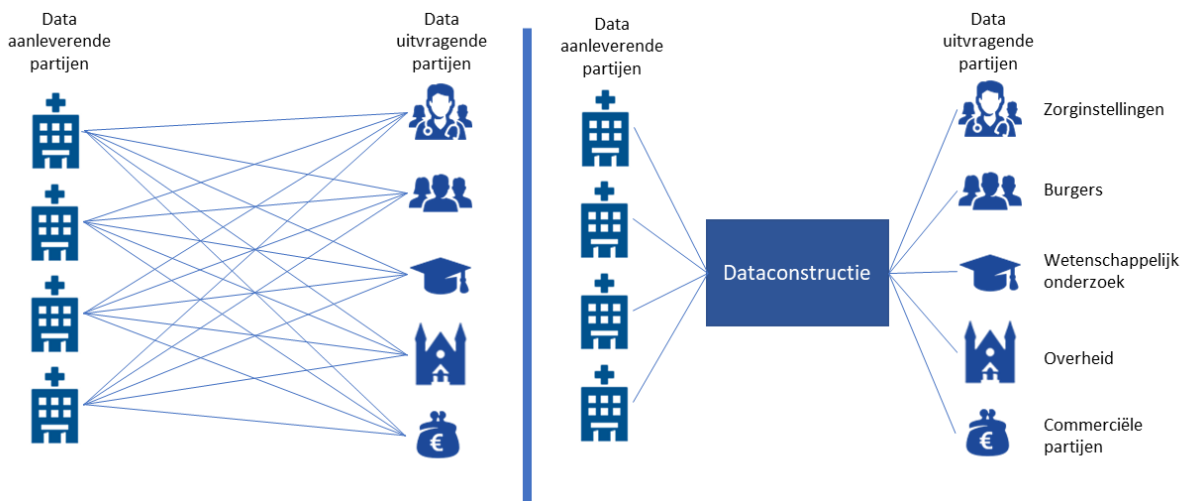
Voor domein van verwerking wordt gekeken in hoeverre dataconstructies de rol van een Trusted Third Party (TTP) innemen. Hierbij is de dataconstructie de poortwachter tussen aanleverende partijen en de uitvragende partijen (zie het voorbeeld in Figuur 4).

Op dit moment wordt veel gebruik gemaakt van centralisatie van data in een dataconstruct. Het dataconstruct verzamelt de data van de aanleverende partijen op één centrale plek (bijvoorbeeld een database) en maakt deze beschikbaar aan de uitvragende partijen. Om gebruik te maken van de data moeten de uitvragende partijen of de data ontvangen van het dataconstruct of toegang krijgen tot de data binnen een beveiligde werkomgeving.

⁶ Datasoevereiniteit houdt in dat de rechthebbende zelf bepaalt met wie en voor wat voor doeleinden de data gedeeld wordt⁷

⁷ Voor een verdieping op de begrippen datasoevereiniteit, vertrouwen en veiligheid, zie: Nederlandse AI Coalitie (NLAIC) (2020). "Verantwoord datadelen voor AI"

⁸ Nederlandse AI Coalitie (NLAIC) (2020). "Verantwoord datadelen voor AI"



Figuur 4 – Datastromen tussen aanleverende en uitvragende partijen zonder en met een beschikbare dataconstructie

Tegenwoordig zijn er ook technieken beschikbaar die de centralisatie van data niet meer vereisen als voorwaarde voor (AI) data-analyse. Bij decentralisatie van data hoeft de data niet meer van plek te verplaatsen. De analyse vindt plaats op de plek van de data. Een voorbeeld hiervan, dat momenteel in Nederland al gebruikt wordt, is de Personal Health Train⁹. De data blijft tijdens de analyses bij de bron en enkel de conclusies brengt de 'trein' terug naar de analist¹⁰. Deze decentrale opslag/beheer van data van data, heeft als grote voordeel dat privacy vaak makkelijker te regelen is. Men voorkomt daarmee het dupliceren van data en de eigenaar houdt meer controle over de data. De keerzijde is dat alle data bij de verschillende bronnen wel op een soortgelijke manier gestandaardiseerd moet zijn voordat analyses succesvol toegepast kunnen worden. Eenheid van taal voor gezondheidsdata is dan ook een grote uitdaging¹¹. Deze uitdaging voor het domein van verwerking moet niet verward worden met bedrijven die veelal worden gebruikt door dataconstructies als vertrouwde derde partij om bijvoorbeeld de data van aanleverende partijen te pseudonimiseren¹².

2. Security & privacy

Datasoevereiniteit wordt door de Nederlandse AI Coalitie expliciet genoemd als uitdaging om data beschikbaar te stellen voor AI binnen de zorg¹³. Gezondheidsdata is inherent gevoelige informatie, omdat het hier om bijzondere persoonsgegevens gaat zoals bijvoorbeeld de fysieke en mentale gezondheid van deze personen¹⁴. Het is wettelijk vereist dat iedereen zelf kan beschikken over zijn eigen data of invloed kan uitoefenen op het gebruik van deze data¹⁵. Binnen de gezondheidszorg is naast datasoevereiniteit het begrip datasolidariteit belangrijk. Met datasolidariteit wordt datadeling van gegevens ten behoeve van het grotere geheel (gezondheid) bedoeld – "De gegevens van de ene patiënt kunnen het leven van een ander verbeteren"¹⁶. De balans vinden tussen datasoevereiniteit en datasolidariteit is een belangrijk vraagstuk voor de verschillende partijen in het veld en vormt daarom een parameter binnen het onderzoek.

⁹ Personal Health Train (2020), <https://pht.health-ri.nl/nl>

¹⁰ Zorginstituut Nederland (2018). "Eindrapportage praktijktoets Personal Health Train"

¹¹ Ministerie van Volksgezondheid, Welzijn en Sport (2018). "Data laten werken voor gezondheid"

¹² ZorgTTP (2020). <https://www.zorgttp.nl/trusted-third-party-ttp/>

¹³ Nederlandse AI Coalitie (NLAIC) (2020). "Verantwoord datadelen voor AI"

¹⁴ Autoriteit Persoonsgegevens (2020). <https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/algemene-informatie-avg/mag-u-persoonsgegevens-verwerken#wat-verstaat-de-avg-onder-bijzondere-persoonsgegevens-6339>

¹⁵ Ministerie van Volksgezondheid, Welzijn en Sport (2018). "Data laten werken voor gezondheid"

¹⁶ Ministerie van Volksgezondheid, Welzijn en Sport (2018). "Data laten werken voor gezondheid"

3. Conformiteit en vertrouwen

Een derde parameter meegenomen in dit onderzoek is conformiteit en vertrouwen. Naast de patiënt/burger, kan ook de beheerder van de gezondheidsdata voorwaarden stellen aan de databeschikbaarheid. De beheerder is namelijk verantwoordelijk voor zijn gezondheidsdata onder zijn beheer en wil misbruik en boetes voorkomen. Daardoor legt de beheerder vaak voorwaarden op over toegang en gebruik van de data. Het uitgangspunt doelbinding is hier van belang, waarbij geverifieerd wordt dat de data enkel en alleen gebruikt wordt voor gerechtvaardigde doeleinde. Ook moet de gebruiker van de data zich conformeren aan de voorwaarden die gesteld worden aan de verwerking van de data. Deze conformiteit zorgt ervoor dat de data verantwoordelijk en betrouwbaar wordt gebruikt. Om ervoor te zorgen dat de databeheerder in controle blijft en de gebruiker altijd verantwoordelijkheid weet af te leggen, is het belangrijk dat de data altijd traceerbaar is¹⁷.

Naast conformiteit is vertrouwen erg belangrijk. Om het vertrouwen van burgers/patiënten in het delen van data voor het algemeen belang is het dus cruciaal dat dit veilig gebeurt, aangezien gezondheidsdata veel (bijzondere) persoonsgegevens bevat. Om het vertrouwen in het gebruik van gezondheidsdata te waarborgen wordt deze data zoveel mogelijk gepseudonimiseerd en geanonimiseerd en ziet wetgeving toe op het gebruik en het naleven van de deze regels.

4. Betrouwbaarheid en kwaliteit

Datakwaliteit is een van de grootste uitdagingen voor de toepassing van AI en een belangrijke variabele tijdens het onderzoek. Om juiste analyses te maken wordt gezondheidsdata uit verschillende bronnen verwerkt. De kwaliteit van deze data kan hierdoor erg verschillen. De kwaliteit van data hangt af van de te ontwikkelen of toe te passen AI, maar doelt in het algemeen op geschiktheid, nauwkeurigheid en volledigheid van deze data. Momenteel is een hoge diversiteit aan verschillende standaarden aanwezig in het veld¹⁸. Dit komt de interoperabiliteit van de data niet ten goede en daarmee de databeschikbaarheid in brede zin. Een voorbeeld van een initiatief om de kwaliteit van data te bewaken ten behoeve van prediction modelling is CHARMS, 'Checklist for critical Appraisal and data extraction for systematic Reviews of prediction modelling Studies'¹⁹. Deze checklist helpt bij het adequater maken van de kwaliteit van verslaglegging bij prediction modelling studies.

3.3 GERELATEERDE ONDERZOEKEN

Data en AI zijn al jaren geen onbekende onderwerpen meer, ook niet in de gezondheidszorg. Demografische ontwikkelingen (vergrijzing), epidemiologische trends (stijging in het aantal mensen met één of meer chronische ziekten), maatschappelijke veranderingen (grotere mondigheid, maar ook groeiende verschillen), technologische innovaties (e-Health, Big Data, personalized medicine) en een andere blik op gezondheid vragen om onderzoek naar een nieuwe, data gedreven aanpak.

De afgelopen jaren hebben onderzoeken een beeld geschetst van de vraagstukken gerelateerd aan databeschikbaarheid in algemene zin en hoe deze kunnen worden opgelost.^{17,20} Daarnaast zijn er een aantal casestudies binnen de gezondheidszorg over succesfactoren van slimme zorg thuis voor patiënten met COPD en hartfalen en zijn op Europees niveau onderzoeken gepubliceerd over een Europese datastrategie, waarbij de huidige problemen omtrent databeschikbaarheid genoemd worden, ook specifiek gericht op het gebruik van data voor AI^{21,22}.

¹⁷ Nederlandse AI Coalitie (NLAIC) (2020). "Verantwoord datadelen voor AI"

¹⁸ Nivel & CBS (2019) "Hergebruik van routine zorgdata voor beleid en wetenschap: hoe het nog beter kan"

¹⁹ Moons et al. (2014) "Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist"

²⁰ Europese Commissie (2020). Witboek over kunstmatige intelligentie - een Europese benadering op basis van excellentie en vertrouwen.

²¹ Zorginstituut Nederland (2018). "Praktijktoets FAIR Data & Personal Health Train"

²² Europese Commissie (2020). "Een Europese datastrategie"

Naast deze onderzoeken die de belangrijkste vraagstukken met betrekking tot databeschikbaarheid uiteenzetten, zijn er ook een aantal onderzoeken gepubliceerd die kijken naar oplossingen voor de toekomst. Een belangrijk voorbeeld hier zijn de internationale FAIR-principes die richting geven aan de manier van beschrijven, opslag en publicatie van wetenschappelijke data. FAIR kent een breed draagvlak in de onderzoekswereld en de voordelen omtrent vindbaarheid, toegankelijkheid, uitwisselbaarheid en herbruikbaarheid worden algemeen onderkend (ook binnen Nederland²³). FAIR maakt het mogelijk verschillende bronnen te koppelen en te interpreteren. FAIR staat voor: Findable (vindbaar); Accessible (toegankelijk); Interoperable (uitwisselbaar); en Reusable (herbruikbaar) en op basis van deze principes kunnen standaarden en tools worden ontwikkeld om de databeschikbaarheid te vergroten.

Uiteindelijk streven al deze onderzoeken naar het verbeteren van de gezondheidszorg in de toekomst. Dit rapport bouwt dan ook voort op de actueel bekende vraagstukken en brengt in kaart wat de overkoepelende thema's zijn om de databeschikbaarheid voor gebruik van AI te vergroten.

4. RESULTATEN

Voor dit onderzoek zijn 63 verschillende organisaties, projecten en initiatieven meegenomen. Van deze organisaties en projecten zijn op basis van de scopingcriteria uiteindelijk 29 constructies geselecteerd die binnen de scope van deze nulmeting vielen. Vanuit deze groep van 29 organisaties en projecten zijn er 23 constructies volledig geïdentificeerd en meegenomen in deze nulmeting. Drie constructies vielen tijdens de fase 'verdiepen' nog af op basis van het interview en de informatie die hier aan het licht kwam. De classificatie van deze verschillende constructies ondersteunt de verdere analyse van de kwalitatieve en kwantitatieve informatie uit de vragenlijsten en interviews. Door middel van deze classificatie is het mogelijk om overkoepelende resultaten tussen de verschillende constructies in kaart te brengen. Daarnaast kunnen per constructie de verschillen en overeenkomsten ten opzichte van elkaar inzichtelijk worden gemaakt.

Met de overzichtsmatrix (Hoofdstuk 4.1) wordt dit doel bereikt door in één oogopslag duidelijk te maken hoe groot de dataverzamelingsscope en het aantal type gebruikers van de verschillende data initiatieven is. In hoofdstuk 5 zijn de resultaten van de onderzoeksgroep verder uiteengezet op basis van de meest relevante en opvallende resultaten. In dit hoofdstuk is een visueel holistisch beeld gegeven van de resultaten uit de onderzoeksgroep, zonder te refereren naar specifieke data constructies. In hoofdstuk 4.3 zijn in detail de verschillende overkoepelende thema's besproken die zijn geïdentificeerd, zoals onder andere de doelgroep (4.3.1), zorgsector (4.3.2) en financiering (4.3.3).

²³ Ministerie van Volksgezondheid, Welzijn en Sport (2018). "Data laten werken voor gezondheid"

4.1 OVERZICHTSMATRIX

De overzichtsmatrix (Figuur 5) is gebruikt om in kaart te brengen hoe de initiatieven verschillen in de scope van hun dataverzameling en hoe zij verschillen in de grootte van de doelgroep die deze data gebruikt. Figuur 5 laat in percentages zien hoeveel constructies in een bepaald cluster vallen. Het gaat hierbij niet om een waardeoordeel te geven over de verschillende clusters ten opzichte van elkaar, maar hier is getracht in kaart te brengen hoe de initiatieven verschillen in de scope van hun dataverzameling en de grootte van de doelgroep die deze data gebruikt.

Deze overzichtsmatrix is opgesteld aan de hand van twee criteria: 1) de gebruikersdoelgroep en 2) de scope van de dataverzameling. Per criteria vallen de verschillende dataconstructies in één van drie 'buckets' (klein, gemiddeld of groot). Door deze twee criteria met ieder 3 opties kunnen constructies in 9 categorieën vallen.

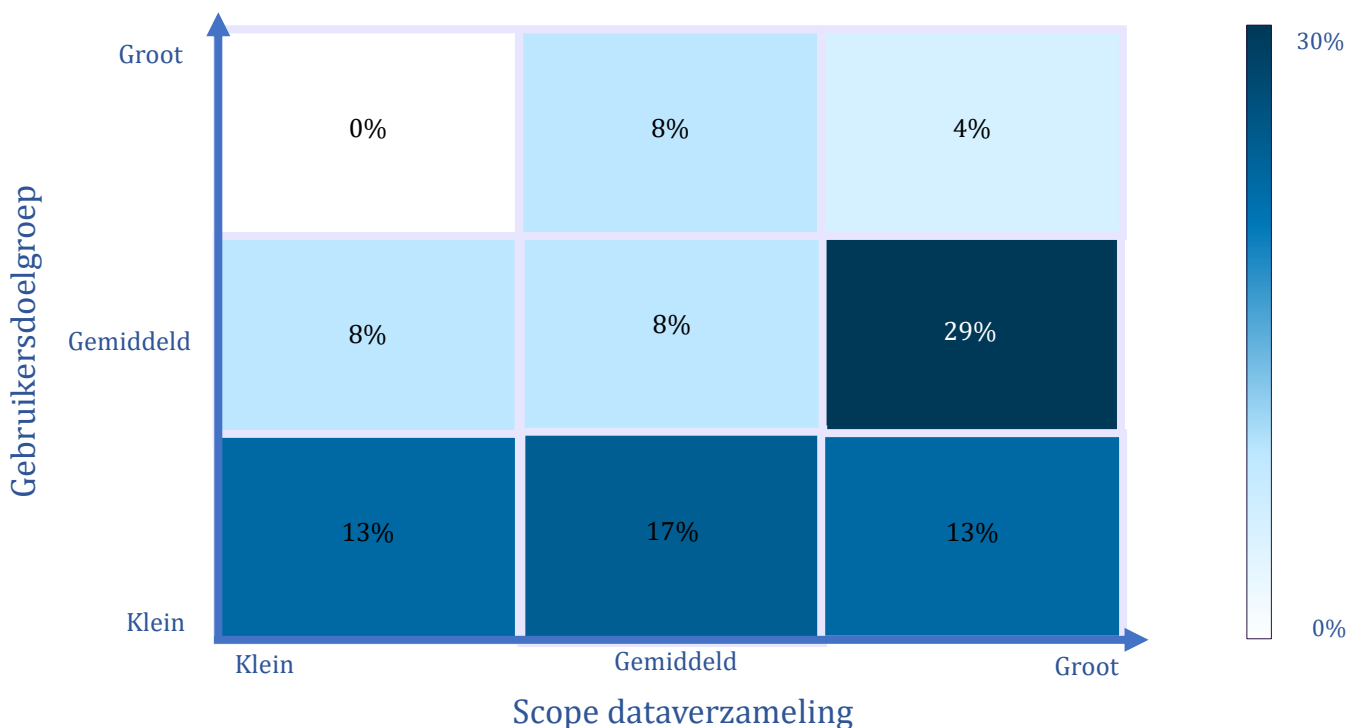
1. Gebruikersdoelgroep

Klein	De dataconstructie richt zich op het beschikbaar maken van data voor één doelgroep
Gemiddeld	De dataconstructie richt zich op het beschikbaar maken van data voor enkele doelgroepen
Groot	De dataconstructie richt zich op het beschikbaar maken van data voor veel verschillende doelgroepen

2. Scope dataverzameling (aandoeningen, ziektebeeld, specialisme)

Klein	De dataconstructie richt zich op het beschikbaar maken van data voor een specifieke medische aandoening, ziektebeeld of medisch specialisme
Gemiddeld	De dataconstructie richt zich op het beschikbaar maken van data voor meerdere specifieke medische aandoeningen met relatie tot elkaar
Groot	De dataconstructie richt zich op het beschikbaar maken van generieke data of meerdere specifieke medische aandoeningen zonder relatie tot elkaar

INDICATIE VAN SCOPE EN BEREIK PER DATA INITIATIEF

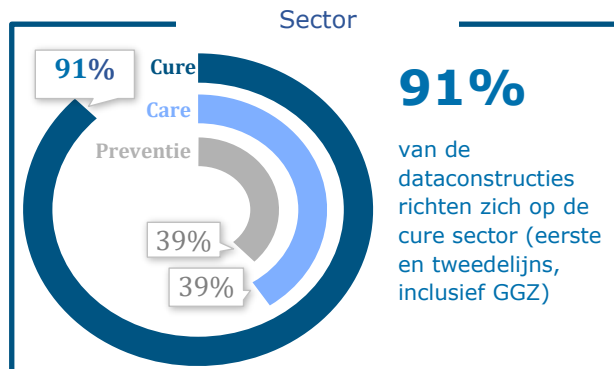


Figuur 5 - Overzichtsmatrix data constructies

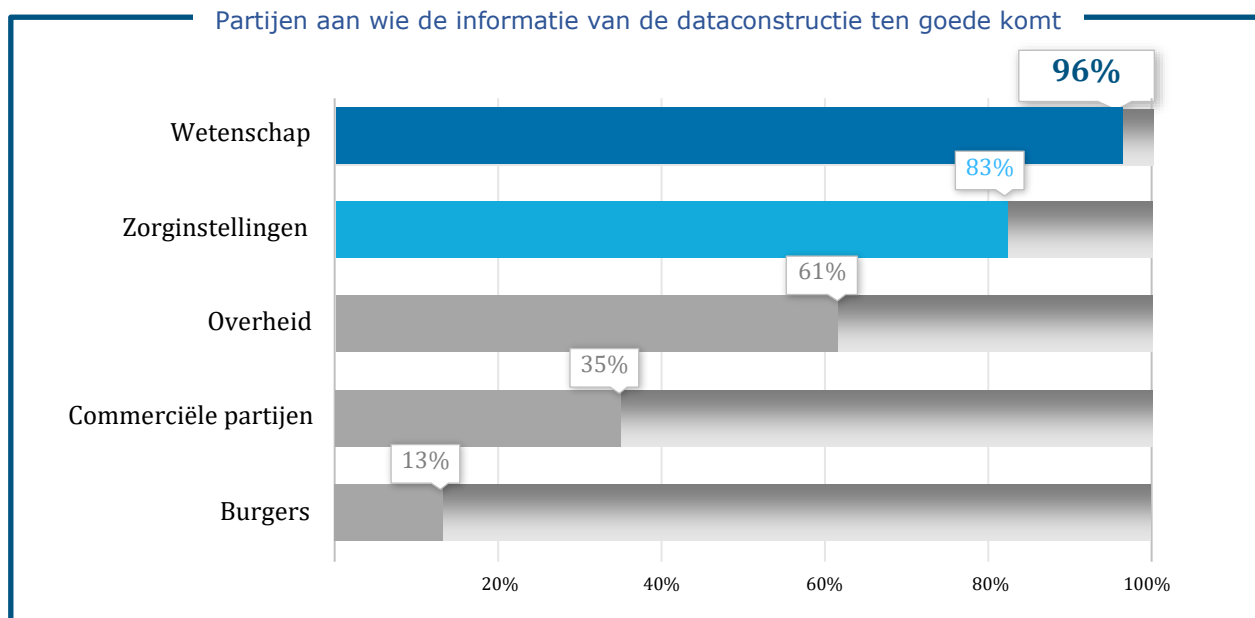
4.2 KENGETALLEN



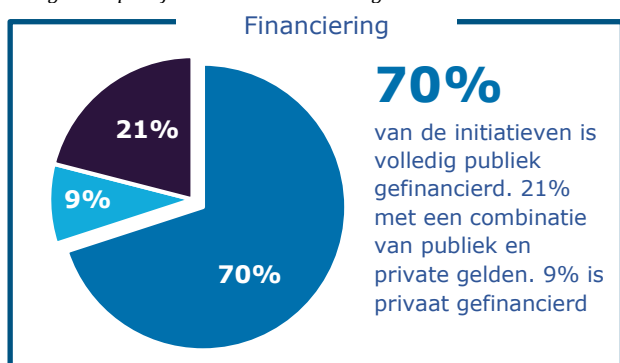
Figuur 6: aantal betrokken initiatieven



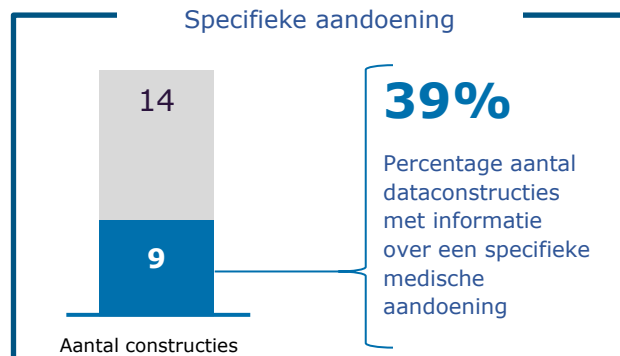
Figuur 7: verschillende zorgsectoren waar constructies zich op richten



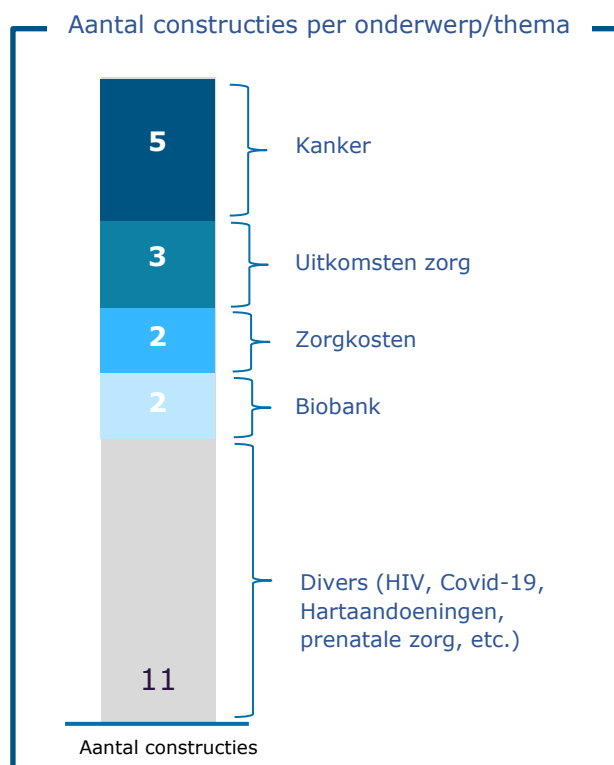
Figuur 8: partijen aan wie de data ten goede komt



Figuur 9: Type financiering per initiatief



Figuur 10: aantal dataconstructies met specifiek ziektebeeld



Figuur 11: aantal dataconstructies per onderwerp/thema

4.3 BESCHRIJVENDE INFORMATIE

De initiatieven bestaan uit een diverse groep organisaties en projecten die vanuit verschillende perspectieven bijdragen aan de toekomst voor gezondheid en zorg in Nederland. Het gaat hierbij om initiatieven die hebben aangegeven mee te willen doen aan het onderzoek. Dataconstructies zijn geen nieuw fenomeen en dragen al decennialang bij aan een betere zorg. Dit blijkt wel uit de oudste constructie die heeft deelgenomen in deze nulmeting en is opgericht in 1950. Daartegenover staan jongere constructies met als oprichtingsjaar de start van het nieuwe decennium, 2020. Daarnaast verschilt het aantal betrokken werknemers per constructie sterk. In deze inventarisatie loopt dit uiteen tussen de 1 en 500 werknemers per organisatie/project. Verder zijn de meeste constructies kennisinstellingen.

Naast deze algemene kenmerken is hieronder een uitsplitsing gemaakt op enkele onderwerpen om toe te lichten wat de belangrijkste verschillen en overeenkomsten zijn tussen de verschillende constructies die zijn meegenomen in deze inventarisatie. Het gaat hierbij om een uitsplitsing naar:

- Doelgroepen voor de databeschikbaarheid (*Hoofdstuk 4.3.1*);
- Zorgsectoren waarin de constructie opereert (*Hoofdstuk 4.3.2*);
- Financieringsstructuren (*Hoofdstuk 4.3.3*);
- Type samenwerkingsmodel voor databeschikbaarheid (*Hoofdstuk 4.3.4*);
- Specifieke aandoeningen (*Hoofdstuk 4.3.5*);
- Relevantie in het vergroten van de waarde voor AI "aan het bed" (*Hoofdstuk 4.3.6*).

4.3.1 DOELGROEPEN

De databeschikbaarheid van de verschillende initiatieven in deze inventarisatie komt ten goede aan verschillende doelgroepen. De geïdentificeerde doelgroepen zijn burgers, de wetenschap (onderzoek), de overheid, zorginstellingen, patiënten en commerciële partijen. Van alle partijen die zijn meegenomen in deze inventarisatie richten 96% zich op het beschikbaar maken van data voor wetenschappelijk onderzoek zoals te zien in Figuur 6. De databeschikbaarheid vergroten voor deze groep resulteert in eenvoudigere dataverzameling door bijvoorbeeld datahergebruik en betere vindbaarheid, standaardisatie van observaties en reproduceerbaarheid van resultaten. Wanneer resultaten gestandaardiseerd en reproduceerbaar zijn kunnen er vaker, sneller en betere analyses worden verricht op grotere datasets wat resulteert in meer en beter onderzoek wat de gezondheidszorg ten goede komt. Een tweede belangrijke bestaande doelgroep voor het beschikbaar maken van gezondheidsdata zijn de zorginstellingen zelf. In totaal maakt 83% van alle initiatieven data beschikbaar voor zorginstellingen. In veel van de gevallen wordt data verzameld omwille van verplichte registraties, maar ook additionele gegevens om inzicht te krijgen in de effectiviteit van specifieke behandelmethodes of het vergelijken van zorgkosten tussen zorginstellingen.

Slechts in 35% van de gevallen in deze inventarisatie wordt data beschikbaar gemaakt aan commerciële partijen (inclusief zorgverzekeraars). Respondenten gaven aan dat de door hen zelf gestelde eisen voor datadeling met partijen met een overkoepelend commercieel doeleinde strenger zijn dan voor partijen zonder winst oogmerk en dat hierdoor datadeling minder vaak voorkomt. Daarnaast is wel benoemd door de dataconstructies dat datadeling met commerciële partijen niet is uitgesloten zolang de privacy gewaarborgd kan blijven en de angst van onrechtmatig gebruik door commerciële partijen uit de lucht is genomen. Ook aan burgers wordt de data op dit moment zeer gelimiteerd beschikbaar gesteld (13%). De data blijft voor deze groep verborgen achter private dataportals en hoge eisen t.a.v. (wetenschappelijk) onderzoek. Voor burgers is er hierdoor beperkt zicht op welke data er precies beschikbaar wordt gemaakt en waar dit voor gebruikt zou kunnen worden. De laatste doelgroep waar naar gekeken wordt is de overheid. Veelal leveren dataconstructies naast bovengenoemde doelgroepen, hun data aan overheidsinstanties zoals het RIVM, NZA of het CBS.



Burgers



Wetenschap



Overheid



Zorginstellingen



Commerciële
partijen

4.3.2 ZORGSECTOREN

In de traditionele verhouding kan de zorg worden ingedeeld in preventie, cure en care. Waar een succesvolle preventie de behoefte aan curatieve zorg vermindert en opvolgend de behoefte aan zorg in het care domein (o.a. verpleging, begeleiding, ondersteuning, verzorging) vermindert. De inventarisatie heeft in kaart gebracht dat 91% van de dataconstructies die op dit moment in het veld bestaan zich onder andere richten op de cure sector. Het merendeel van de data voor deze constructies komt voornamelijk vanuit (academische) ziekenhuizen en onderschrijft het inzicht in sectie 4.3.1 dat databeschikbaarheid voornamelijk ten goede komt aan (wetenschappelijk) onderzoek. Voor constructies die zich richten op de preventie (39%) en care (39%) sector is het complexer om data te verzamelen, omdat deze data breder verspreid is door Nederland en vaak in kleinere organisaties wordt bijgehouden en verwerkt. Daarbij geldt ook voor deze constructies dat hun datasets vaker kleiner zijn en specifiek zijn toegespitst op een bepaalde aandoening of thema.

4.3.3 FINANCIERING

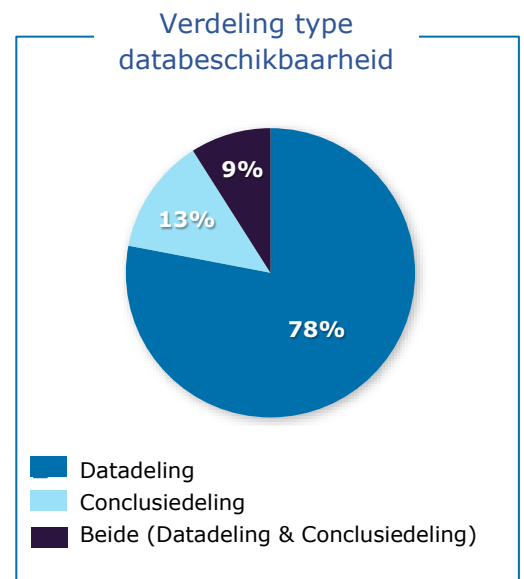
De financiering van de verschillende dataconstructies komt merendeels vanuit uitsluitend publieke gelden (70%). Een kleiner aantal constructies krijgt naast een publiek geld ook een klein deel privaat geld, maar dit is de minderheid. Dataconstructies die worden betaald door (academische) ziekenhuizen worden gerekend onder publieke financiering. Van deze groep geeft 71% aan dat financiering de schaalbaarheid en toekomst van de constructie in de weg staat. In enkele gevallen (3/23) is geen losse financiering voor het initiatief beschikbaar en wordt het initiatief betaald via het standaard salaris van deze werknemer die uren uit zijn dag neemt om dit initiatief op te bouwen en onder de aandacht te brengen bij vakgenoten.

4.3.4 SAMENWERKINGSMODELLEN

Ieder initiatief draagt op een andere manier bij aan de databeschikbaarheid van gezondheidsdata voor AI die uiteindelijk kan leiden tot AI die de gezondheid van patiënten en burgers in Nederland kan ondersteunen. Het gaat hier bijvoorbeeld over het beschikbaar maken van datasets van protontherapie voor onderzoek, het organiseren en bruikbaar maken van alle pathologie informatie voor patiëntenzorg en wetenschap of het bij elkaar brengen van data over de langdurige zorg. De dataconstructies in dit onderzoek zijn de schakel tussen de brondata en de uitvragende partijen. Hoe de data beschikbaar wordt gemaakt vanuit de dataconstructie aan de afnemende partijen gebeurt op hoofdlijnen nog altijd op twee verschillende manieren:

1. De data gaat naar de plek van de analyse (**Datadeling**)
 - De (ruwe) data wordt gedeeld met de uitvragende partij. De analyse wordt uitgevoerd door de uitvragende partij los van de organisatie die de data beschikbaar heeft gemaakt.
2. De analyse gaat naar de plek van de data (**Conclusiedeling**)
 - De data wordt niet gedeeld met de uitvragende partij en de analyse wordt toegepast op de data op locatie van de organisatie die de data beschikbaar maakt.

Er zijn constructies die beide vormen toepassen om de data beschikbaar te maken. In veel gevallen wordt hierbij gekeken naar het verzoek van de uitvragende partij en geïnventariseerd op welke manier aan dit data beschikbaarheidsvraagstuk kan worden voldaan.



4.3.5 SPECIFIEKE AANDOENINGEN

In deze inventarisatie is ook gekeken of de verschillende initiatieven data verzamelen of beheren van specifieke aandoeningen of medische specialismen. Dit is meegenomen in dit onderzoek om inzichtelijk te maken in hoeverre initiatieven worden opgezet om gezondheidsdata rond een specifiek onderwerp op te zetten. Uit deze inventarisatie blijkt dat 61% van de initiatieven zich niet richt op een specifieke aandoening of medische specialisme. Een mogelijke verklaring hiervoor is dat naarmate constructen langer bestaan ze de mogelijkheid krijgen om te groeien en meer data te ontsluiten over verschillende aandoeningen of medisch specialisme heen. Dit komt ook terug uit de verdiepende gesprekken waar sommige respondenten aangeven meer data te willen ontsluiten over een breder scala aan aandoeningen en specialisme. Dit is niet altijd het geval aangezien in deze inventarisatie ook constructen zijn meegenomen die al sinds de jaren 70/80 bestaan en zich nog altijd toeleggen op een specifiek medisch specialisme.

4.3.6 WAARDE VAN DATA VOOR AI "AAN HET BED"

Tenslotte is binnen dit onderzoek gekeken naar de waarde van data die het mogelijk maakt om AI toe te passen bij de directe zorgverlening van patiënten, zogenaamde AI "aan het bed". Voorbeelden hiervan zijn zorgapparatuur of informatiesystemen die gebruik maken van AI om aandoeningen te diagnosticeren of artsen te adviseren over behandelmethodes op basis van de medische gegevens van de patiënt. In deze inventarisatie is gekeken naar in hoeverre de constructen bijdragen aan de databeschikbaarheid van data die relevant is voor deze AI toepassingen met een directe impact op de zorgverlening aan een patiënt. Om onderscheid aan te brengen in de relevantie van constructen voor AI toepassingen is de volgende classificering gehanteerd.

Waarde van data voor AI "aan het bed"	
<u>Zeer relevant</u>	De dataconstructie beschikt over data zoals diagnoses, behandelmethodes, uitkomsten zorg, opgenomen beelden of monsters op patiëntniveau
<u>Relevant</u>	De dataconstructie beschikt over data zoals ligduur, opnames, declaraties en kosten
<u>Beperkt relevant</u>	De dataconstructie beschikt over data vanuit landelijke gemiddeldes of geaggregeerde datasets

Uit de inventarisatie blijkt dat 61% van de constructies binnen de categorie 'Zeer relevant' vallen. Dit betekent dat de data van deze dataconstructies te gebruiken is voor AI toepassingen die een bijdrage kunnen leveren aan de zorgverlening van een patiënt "aan het bed". 4% van de dataconstructies in deze inventarisatie verzamelt data die slechts beperkt relevant is om de waarde van AI toepassingen aan het bed van patiënten kan vergroten. Het gaat hierbij om data die voornamelijk wordt gebruikt voor benchmarks en rapporten die op landelijk of regionaal niveau worden gepresenteerd.

4.4 DATAVRAAGSTUKKEN

De verschillende constructies in deze inventarisatie hebben allemaal te maken met datavraagstukken die relevant zijn binnen het zorgdomein. Zoals benoemd in hoofdstuk 3 zijn de belangrijkste datavraagstukken voor de zorg op te delen in vier verschillende categorieën:

1. Domein van verwerking
2. Security & Privacy
3. Conformiteit en vertrouwen
4. Betrouwbaarheid en kwaliteit

Ieder initiatief heeft specifieke datavraagstukken die gelieerd kunnen zijn aan één van bovenstaande categorieën. De antwoorden vanuit de verschillende initiatieven rondom deze datavraagstukken zijn zeer divers, daarom worden per categorie de belangrijkste en meest gehoorde datavraagstukken toegelicht.

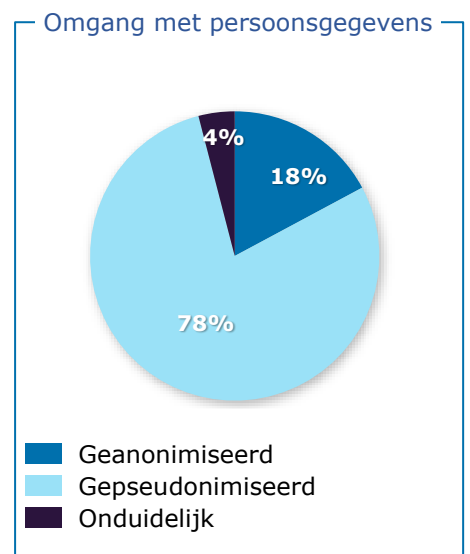
4.4.1 DOMEIN VAN VERWERKING

Data in de vorm van opgenomen signalen of afbeeldingen (bijv. CT-en MRI-scans) zijn grote bestanden in vergelijking met numerieke of tekstuele data. Voorafgaand aan dit onderzoek werd dan ook aangenomen dat dit een potentieel pijnpunt is voor veel dataconstructies. In de praktijk is dit niet het geval en ervaren veel initiatieven de data opslag van dit type data niet als een groot vraagstuk. De techniek is hierbij niet de belemmerende factor, onderliggende vraagstukken zoals ontoereikende financiering om de dataopslag uit te breiden zijn dat wel.

Ten aanzien van het domein van verwerking zijn er wel verschillen te benoemen in de aanpak tussen de verschillende constructies. In het veld zijn op dit moment twee denkwijze hoe om te gaan met het opslag en beschikbaarheid van gezondheidsdata. Het verschil zit hem hier in het of centraliseren van zorggegevens of het gedecentraliseerd houden van informatie. Waar sommige initiatieven inzetten op het vergroten van hun constructies door nieuwe bronnen te ontsluiten, kiezen andere constructies ervoor om samen te werken met andere constructies. Samenwerking kan bestaan uit het koppelen van datasets, maar ook het delen van conclusies op onderzoeksvragen vanuit dataconstructies. Doormiddel van samenwerking kunnen dataconstructies zich specialiseren in onderwerpen en thema's, maar toch toegang blijven behouden tot data die wordt ontsloten door andere partijen. Verdere centralisatie van data bij één partij wordt in het veld ook niet als wenselijk gezien en dit blijkt uit het toekomstperspectief van het veld om meer in te zetten op samenwerking.

4.4.2 SECURITY & PRIVACY

Alle constructies in deze inventarisatie geven aan te voldoen aan alle wetgeving met betrekking tot het beheer en/of verwerking van gezondheidsdata. Al dan niet doordat zij aangeven te werken met geanonimiseerde data. Geanonimiseerde gezondheidsdata worden zodoende niet langer als persoonsgegevens beschouwd²⁴, behalve als er een grote kans is dat technologieën het mogelijk maken de data weer te her-identificeren. De opkomst van AI heeft deze kans vergroot doordat zeer grote hoeveelheden data geanalyseerd kunnen worden en patronen makkelijker herkend worden. AI brengt hier een risico met zich mee dat doordat dit soort toepassingen in staat zijn om geanonimiseerde data dusdanig te koppelen aan andere bronnen dat deze data alsnog te herleiden valt naar een identificeerbaar persoon. Is dat het geval spreken we niet meer over geanonimiseerde gezondheidsdata maar over persoonsgegevens. De gesproken constructen in dit onderzoek zijn zich bewust van dit risico en geven aan dat dit risico geminimaliseerd is. Daarom worden



²⁴ Autoriteit persoonsgegevens (2020), <https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/avg-europese-privacywetgeving>

geanonimiseerde data in dit onderzoek niet als persoonsgegevens beschouwd. Voor persoonsgegevens die zijn gepseudonimiseerd geldt dat deze zijn versleuteld en de koppeling naar een natuurlijk persoon alleen gemaakt kan worden door aanvullende gegevens (de sleutel) die apart bewaard dienen te worden. Gepseudonimiseerde gezondheidsdata worden dus wel beschouwd als persoonsgegevens. Enkele respondenten geven aan dat anonimisering en/of pseudonimisering leidt tot dataverlies, waarbij zij de directe koppeling met de patiënt niet meer kunnen maken. Deze koppeling van data op basis van patiëntgegevens is een tweede veel gehoord datavraagstuk voor veel constructies. Doordat deze directe koppeling (bijvoorbeeld op BSN) op dit moment een wettelijke grondslag vereist, geven de respondenten aan dat potentiële waarde verloren gaat doordat samenwerking met andere constructies inefficiënter is en doordat afnemers van deze data zelf de data gaan koppelen op basis van de gemakkelijkst beschikbare dataset in plaats van de meest geschikte dataset. Naar aanleiding van het commissierapport van der Zande²⁵ is het ministerie van VWS op dit onderdeel bezig met het onderzoeken van een wettelijke grondslag die voor kwaliteitsregistraties kan worden gehanteerd.

4.4.3 CONFORMITEIT EN VERTROUWEN

De aanleverende partijen leveren hun data vrijwillig dan wel verplicht aan bij de verschillende dataconstructies in deze inventarisatie. In veel gevallen houden de aanleverende partijen het recht om de data die zij aanleveren in te zien en te gebruiken zonder restricties. Hiernaast kunnen aanleverende partijen ervoor kiezen, bij niet verplichte datadeling, om hun data terug te halen en niet langer beschikbaar te maken via de dataconstructie. Voor de uitvragende partijen die gebruik willen van de data moet in alle gevallen de gebruiker zich conformeren aan de voorwaarden die gesteld worden door de constructen. Alleen waar het openbare data betreft heeft iedereen vrije toegang tot de data (bijvoorbeeld de databanken van het CBS²⁶, het RIVM²⁷, de NZA²⁸)

Volgens een rapport van de Patiënten Federatie Nederland over patiënten waarbij lichaamsmateriaal wordt afgenomen is in 68% van de gevallen niet om toestemming gevraagd voor het gebruik van medische gegevens voor wetenschappelijk onderzoek. Het overgrote deel van patiënten, indien zij dit gevraagd zou worden, geeft aan geen moeite te hebben het delen van gezondheidsdata voor onderzoek²⁹. Het patiëntconsent blijft een uitdaging voor veel constructies. Dataconstructies zouden graag met terugwerkende kracht consent verzamelen of een patiënt informeren over uitkomsten, maar een patiënt traceren vanuit de data is juridisch gezien niet mogelijk. Dit brengt ook vanuit ethiek een uitdaging met zich mee, omdat dataconstructies een patiënt graag op de hoogte zou willen houden van wat er met zijn data gebeurt na de consentverlening. Zo zouden verschillende constructen graag de mogelijkheid hebben om in de toekomst patiënten te kunnen contacteren mochten er met nieuwe technieken bijvoorbeeld aandoeningen gevonden worden op basis van eerder gedeelde informatie.

4.4.4 BETROUWBAARHEID EN KWALITEIT

De betrouwbaarheid en kwaliteit van data is van essentieel belang om de waarde van data te kunnen benutten, zeker voor AI-toepassingen. Doordat veel gezondheidsdata ontstaat binnen verschillende organisatie en lokale zorginstellingen is de kwaliteit van de data divers waardoor het moeilijker is om verschillende bronnen met elkaar te combineren. Vanwege deze verschillen in dataregistratie geven vrijwel alle constructies in deze inventarisatie aan dat zij worstelen met de heterogeniteit van de data die zij ontvangen. Om dit probleem op te lossen stimuleren dataconstructies standaardisatie van data bij de bron door middel van procedures en bieden enkele constructies tools aan om de zorginstellingen hierbij te helpen. Deze initiatieven om de kwaliteit van de data te verbeteren worden veelal gestart vanuit de dataconstructies zelf.

²⁵ Advies commissie governance van kwaliteitsregistraties (2019), <https://www.rijksoverheid.nl/documenten/rapporten/2019/07/05/advies-commissie-governance-van-kwaliteitsregistraties>

²⁶ CBS, CBS Statline (2020) <https://opendata.cbs.nl/#/CBS/nl/>

²⁷ RIVM, RIVM Statline (2020) <https://statline.rivm.nl/#/RIVM/nl/>

²⁸ NZA, DIS (2020) <https://www.opendisdata.nl/dis/over>

²⁹ Patiënten Federatie Nederland (2019), <https://www.patiëntenfederatie.nl/actueel/nieuws/delen-medische-data-voor-meeste-patienten-geen-punt>

4.5 VOORWAARDEN VOOR DATABESCHIKBAARHEID

Met huidige aanpak vanuit het veld om de databeschikbaarheid van gezondheidsdata te vergroten komen vraagstukken omhoog over welke aanvullingen nodig zijn om gebruik van AI voor gezondheid en zorg te stimuleren. Om dit te identificeren zijn de verschillende constructen gevraagd naar welke obstakels de schaalbaarheid van deze initiatieven in de weg liggen volgens de respondenten zelf. Wat zijn voor hen de belangrijkste voorwaarden om de complementariteit met andere initiatieven te vergroten en wat is volgens de initiatiefhouders nodig is om in de toekomst gebruik te kunnen maken van AI? De antwoorden op deze vragen kunnen worden samengevat in de drie thema's die vanuit de dataconstructies handvatten moeten bieden om de databeschikbaarheid voor AI van gezondheidsdata te vergroten. Dit betreft de volgende drie thema's welke onderstaand zijn toegelicht:

- Regelgeving (4.5.1)
- Harmonisatie (4.5.2)
- Vertrouwen (4.5.3)

4.5.1 REGELGEVING

Vanuit het veld worden twee belangrijke obstakels benoemd die de waarde van gezondheidsdata kunnen vergroten en kunnen leiden tot meer gebruik van AI, maar die door huidige regelgeving zonder expliciete toestemming niet mogelijk zijn. Het gaat hierbij om de wens om data te kunnen koppelen op patiëntniveau en meer ruimte ten aanzien van doelbinding. Binnen de huidige kaders is dit niet toegestaan vanwege privacy redenen en om misbruik te voorkomen, maar dientengevolge leidt dit tot beperktere mogelijkheden om data te gebruiken voor AI binnen de zorg. Daarnaast wordt vaak benoemd dat interpretatie van de AVG als lastig wordt ervaren, omdat iedere organisatie deze zorgregels op een andere/eigen manier interpreteert en uitlegt. Hierdoor zijn partijen terughoudend in het delen van data uit angst om op de vingers getikt te worden. Terughoudendheid is niet per definitie verkeerd, maar het beperkt de snelheid van innovatie op dit gebied. Vooral de grote partijen in deze inventarisatie geven aan dat door rigide interpretatie van wetgeving het grootschalig data verzamelen en delen te veel wordt beperkt. Voor deze partijen is het beginsel van 'doelbinding'³⁰ ook het meest relevant aangezien deze partijen de behoefte en data hebben om met big-data analyse patronen te analyseren in hun datasets. Daarnaast zouden partijen graag data willen koppelen uit verschillende datasets binnen of buiten de dataconstructie. Het is echter niet toegestaan om direct de data te koppelen op basis van een individu. Er zijn mogelijkheden om deze data te combineren via een waarschijnlijkheidskoppeling. Hierbij wordt data gekoppeld op bijvoorbeeld geboortjaar, woonplaats en geslacht. Echter, komt hierdoor mogelijk wel de anonimiteit van persoonsgegevens in de dataset in gevaar. Hierdoor kunnen de gegevens te herleiden zijn tot een identificeerbaar persoon wat alleen is toegestaan bij wettelijk uitzondering. Als oplossing hiervoor gaven meerdere constructies aan dat een koppeling op basis van het geëncrypt BSN een mogelijkheid is. Zij gaan ervan uit dat dit minder privacy belastend is, omdat voor koppeling op het BSN slechts één veld nodig is om data te kunnen koppelen. Daarnaast zou deze koppeling op geëncrypt BSN ook minder tijd en geld kosten, omdat deze efficiënter te maken is. Dit in tegenstelling tot een waarschijnlijkheidskoppeling waarbij gebruik moet worden gemaakt van verschillende persoonlijke data zoals leeftijd, postcode of geslacht om dezelfde koppeling te kunnen maken. Hiertegenover staat dat koppeling van data op basis van een geëncrypt BSN koppelingen dusdanig vereenvoudigen dat de doelmatigheid van analyses in het gedrang kunnen komen. De huidige tijdrovende manier van datakoppelen zorgt er namelijk wel voor dat deze koppelingen niet te pas en te onpas worden gemaakt. Het onderwerp koppeling op basis van BSN is reeds door verschillende partijen onder de aandacht gebracht van VWS en wordt nader onderzocht.

³⁰ <https://autoriteitpersoonsgegevens.nl/nl/over-privacy/persoonsgegevens/verstrekken-van-persoonsgegevens>

4.5.2 HARMONISATIE

Het gezondheidsdata landschap in Nederland is divers maar ook versnipperd. Doordat de dataconstructies zich richten op specifieke onderwerpen, data of aandoeeningen bevindt gezondheidsdata zich op veel verschillende plekken. Vanuit het oogpunt van databeschikbaarheid is dit suboptimaal, maar hierdoor kunnen constructies zich wel specialiseren op specifieke onderwerpen en daardoor beter toezien op de kwaliteit, volledigheid en betrouwbaarheid van de data. Een balans moet hier worden gezocht tussen databeschikbaarheid en datakwaliteit, waarbij een oplossing voor de fragmentatie van gezondheidsdata mogelijk is door datacentralisatie of vergrote samenwerking tussen de verschillende constructies. Datacentralisatie over huidige dataconstructies heen wordt door bijna alle partijen als ongewenst en onrealistisch beschouwd, onder andere vanwege de beperktere mogelijkheden om toe te zien op datakwaliteit. De oplossing moet komen vanuit een bredere samenwerking tussen de verschillende constructies in het veld om gezamenlijk bij te dragen aan de databeschikbaarheid. Het grootste obstakel voor samenwerking is dat er veel verschillende initiatieven (al dan niet tijdelijk) in het veld zijn en dat deze partijen nog vaak gebruik maken van verschillende standaarden en structuren voor data. In het veld wordt dan ook niet gesproken over een gebrek aan data, maar over een gebrek aan toegankelijke data. De FAIR-principes³¹, die ten grondslag zouden moeten liggen aan gezondheidsdata, komen hierdoor in gevaar. De versnippering van standaarden zorgt ervoor dat data tussen constructies lastiger kan worden gedeeld (*Interoperable*). Daarnaast vraagt het meer werk van de aanleverende partij om aan deze standaarden per constructie te voldoen. Bovendien maakt de versnippering van constructies in het veld het moeilijker voor uitvragende partijen om te vinden waar welke data aanwezig is en is daarbovenop ook ongunstig voor de overheid om bij te houden wat de bijdrage van deze initiatieven is aan de databeschikbaarheid binnen de zorg (*Findable*). Op dit moment zijn er verschillende gemandateerde uitvoeringsorganisaties zoals ZonMW, het RIVM, maar ook Health-Holland en Health-RI en technische organisaties (o.a. TNO, GO-FAIR, DTL, SURF) die dit probleem proberen op te lossen. Dit gebeurt enerzijds door alleen subsidies of financiering te verlenen aan initiatieven als er een gebudgetteerd data stewardship plan aanwezig is inclusief FAIR data en services aspecten. Anderzijds, ondersteunen deze partijen door als landingsplek te fungeren voor data initiatieven (bijv. [Health-RI](#)) of technische kennis en tooling te delen hoe gezondheidsdata volgens de juiste procedures beschikbaar te maken (bijv. [GO-FAIR](#)).

Deze lijn past ook binnen de toekomstvisie van de verschillende constructies die zijn meegenomen in deze inventarisatie. Zij geven aan dat ze willen toezien op een bredere databeschikbaarheid door beter vindbaar en toegankelijk te zijn zodat een grotere groep gebruik kan maken van de constructie. Daarnaast is de wens om complementariteit te vergroten tussen de verschillende initiatieven en dit is alleen mogelijk als de interoperabiliteit wordt verbeterd en er tevens een focus is op het hergebruik van deze data.

4.5.3 VERTROUWEN

Zoals benoemd in hoofdstuk 3 ligt vertrouwen ten grondslag aan betere databeschikbaarheid. Zonder dit vertrouwen zijn patiënten en burgers minder snel geneigd om hun gezondheidsdata te delen. Dit is begrijpelijk vanuit de context dat een burger of patiënt zelf ook geen vrije beschikking heeft over zijn eigen zorggegevens. De vraag komt dan op waarom hij dit wel zou willen delen. Op dit moment geven verschillende dataconstructies zelf aan dat er geen terugkoppeling is naar patiënten waarvoor hun gegevens worden gebruikt. Hier ligt een kans voor deze partijen om een voorbeeld te zetten en aan burgers en patiënten terug te koppelen voor welke onderzoeken hun gegevens worden gebruikt en wat de uitkomsten hiervan zijn. Daarnaast hebben dataconstructies, zorgverleners en onderzoekers zelf de verplichting om hun patiënten goed te informeren over het belang van datadeling ten behoeve van datasolidariteit. De verwachting is dat dit de acceptatie voor gezondheidsdatadeling verhoogt als de toegevoegde waarde helder is. Om breder gebruik van AI mogelijk te maken stellen deze

³¹ Zie Hoofdstuk 2.33.3 voor een uitleg over de FAIR principes

constructies ook voor om vanuit de overheid stappen te zetten om publieke vertrouwen in het gebruik van data te vergroten en te tonen welke inzichten en toepassingen het Nederland kan brengen. Het programma Waardevolle AI voor gezondheid is een voorbeeld van hoe de overheid bijdraagt aan het creëren van dit vertrouwen onder organisaties, patiënten, zorginstellingen en burgers.

Daarnaast is onzekerheid over financiering een vaker gehoord probleem. Niet alle dataconstructies hebben een stabiele financieringsbasis waardoor zij moeite hebben om een lange termijn visie uit te zetten. Dit belemmert deze initiatieven om verder te groeien doordat ze niet kunnen investeren in nieuwe technieken of de mankracht hebben om in te zetten op het gebruik van AI of nieuwe samenwerkingsverbanden aan te gaan. Deze constructies geven aan dat een duidelijke financieringsstrategie met een langere horizon de stabiliteit kan bieden waarop zij structuur kunnen aanbrengen om te werken aan opschaling of verbetering van de initiatieven. Zij benoemen hier dat zij vanuit de overheid een rol zien om het vertrouwen in deze dataconstructies uit te stralen. Veel dataconstructies geven hierbij aan dat ze enkel gefinancierd worden vanuit het oogpunt van onderzoek (bijvoorbeeld NWO of ZonMw subsidies). Zij missen hier het vertrouwen en de steun vanuit de zorg en overheid om ook bij te dragen aan de financiering om zo de databeschikbaarheid te vergroten. Aanspraak maken op lange termijn financiering maakt het voor deze constructies mogelijk om lange termijn doelen te stellen waarop de zorg en overheid invloed uit kan oefenen. Daarnaast is tijdens deze inventarisatie vaak de hulpvraag gekomen vanuit het veld om meer en duidelijkere organisatorische sturing vanuit de overheid. Naast een additionele verantwoordelijkheid en rol is dit ook een handreiking vanuit het veld om meer samen te willen werken met de overheid. Er is volgens verschillende partijen een rol weggelegd voor de overheid om sturing te geven aan het veld door een eenvoudig te hanteren methode voor uitwisseling en codering van zorginformatie op basis van datastandaarden (zoals bijvoorbeeld SNOMED en HL7 FHIR) uit te zetten al dan niet op te leggen. Het inzetten op de brede adoptie van FAIR principes wordt ook benoemd en erkend door de overheid. Zij zien een grotere rol voor de overheid om adoptie van deze principes te koppelen aan de financiering van nieuwe en bestaande initiatieven naar voorbeeld van ZonMw³². Dit schept vertrouwen in het gehele gezondheidsdomein wetende dat met gedeelde afspraken het veld zich dezelfde kant op ontwikkeld. Daarnaast moeten constructies beoordelen of zij ook financiering kunnen aantrekken vanuit de private sector. Organisaties zoals Health Holland³³ stimuleren multidisciplinaire publiek-private samenwerkingsverbanden waar deze dataconstructies aanspraak op zouden kunnen maken, maar alleen als de dataconstructie het vertrouwen kan scheppen dat zij daadwerkelijk waarde toevoegen binnen hun domein. Hiervoor moeten de verschillende initiatieven zelf uitwerken op welke manier zij bijdragen aan de databeschikbaarheid en hiermee hun toegevoegde waarde bewijzen ten aanzien van de ambities voor een meer data-gedreven zorg.

³² ZonMW, FAIR data and data management (2020), <https://www.zonmw.nl/en/research-and-results/fair-data-and-data-management/>

³³ <https://www.health-holland.com/about-us>

5. CONCLUSIE & AANBEVELINGEN

Uit de inventarisatie onder dataconstructies die bijdragen aan de databeschikbaarheid voor gebruik van AI in gezondheid en zorg komt naar voren dat er geen gebrek is aan gezondheidsdata, maar wel aan toegankelijke gezondheidsdata. Momenteel ligt daarom de focus in dit veld vooral op het verzamelen, verwerken/bewerken en beschikbaar stellen van data via samenwerkingsconstructies of het ontsluiten van additionele bronnen. Daarnaast wordt data en kennis gedeeld met verschillende dataconstructies die andere onderwerpen, aandoeningen of thema's adresseren. Op dit moment is de algehele tendens in het veld om waar mogelijk meer data te ontsluiten bij verschillende bronnen en om deze vindbaar te maken via catalogussen en portals om ervoor te zorgen dat deze data benaderd kan worden door onderzoekers en AI-toepassingen in de toekomst. Het vindbaar en toegankelijk maken van deze informatie wordt voor veel constructies als een eerste stap gezien. Waar enkele dataconstructies aangeven zelf meer gezondheidsdata te willen verzamelen is de verwachting niet dat verdere centralisatie van data de toekomst is binnen het gezondheidsdomein. Vanwege de huidige regelgeving die toeziet op zorgvuldig en doelbewust delen van gezondheidsdata, maar ook initiatieven waarmee decentraal opgeslagen data kan worden gebruikt (bijvoorbeeld de Personal Health Train³⁴) is het niet nodig om data in toenemende mate centraal te verzamelen. Daarnaast zijn er bezwaren op centralisatie van gezondheidsdata gelet op het vertrouwen (privacy) en veiligheid (cyber security) die hiermee in het geding komen.

Voor veel partijen in het veld is het vindbaar en toegankelijk maken van gezondheidsdata op dit moment het hoofddoel. De toegevoegde waarde van gezondheidsdata voor toepassingen in de praktijk voor patiënten en zorgverleners wordt nog beperkt onderzocht. Juist de invalshoek om te kijken naar toepassingen in de praktijk en de waarde die dit voor mens en zorg kan toevoegen zouden er weer toe kunnen leiden dat juist data makkelijker beschikbaar wordt gemaakt. De databeschikbaarheid komt nu vooral nog (wetenschappelijk) onderzoek, verplichte registraties en rapportages ten goede. Het bijdragen aan databeschikbaarheid voor de ontwikkeling of toepassing van AI is niet het hoofddoel. Op dit moment is dit slechts een positieve bijkomstigheid en/of resultante van de keuzes die in het veld worden gemaakt. AI wordt op dit moment in het veld gezien als de volgende stap na brede databeschikbaarheid van gezondheidsdata. De databeschikbaarheid die wordt geboden door deze constructies kan worden gebruikt voor AI, maar vergt in veel gevallen een lange termijn visie voordat AI daadwerkelijk ontwikkeld en toegepast kan worden met deze data. Daarnaast zijn op dit moment verschillende partijen vaak nog op zoek naar hoe AI het beste toegepast zou kunnen worden zodat het ook waarde kan toevoegen voor hun onderwerp, thema of segment.

Deze conclusies leiden tot de vraag welke bruikbare oplossingen er zijn om databeschikbaarheid voor gezondheid en zorg in Nederland te borgen. Onderstaand worden drie handelingsperspectieven geboden die bijdragen aan het vergroten van de databeschikbaarheid en die het gebruik van AI in het veld ondersteunen.

5.1 SAMENBRENGEN VAN INITIATIEVEN

Het huidige landschap van dataconstructies is zeer gefragmenteerd. Fragmentatie in het veld leidt tot gelimiteerde vindbaarheid, toegankelijkheid en interoperabiliteit van data. Veel dataconstructies zijn vanuit hun domein bezig met het beschikbaar maken van data relevant voor enkel hun directe doelgroep. Hierdoor ontstaan talloze initiatieven die verschillende standaarden en procedures hanteren en nog niet bezig zijn met de vraag hoe hun data gebruikt kan worden door andere personen, organisaties of AI-toepassingen dan hun initiële doelgroep. Het risico dat hierbij ontstaat is dat deze manier van werken op de lange termijn de databeschikbaarheid kan vertragen. Vanuit het veld ontstaan er platformen en (koepel)organisaties die streven om deze fragmentatie te verkleinen. Deze organisaties vervullen een belangrijke functie, omdat zij initiatieven aan zich binden, de manier van

³⁴ Personal Health Train (2020), <https://pht.health-ri.nl/nl>

werken harmoniseren en het pad naar de toekomst uitstippelen om de gezondheidszorg in Nederland samen te verbeteren.

Het is niet alleen voor dataconstructies, maar ook voor de overheid van groot belang om aan te sluiten bij deze initiatieven. Door actief samen te werken met het veld kan er worden meegedacht over de richting en het einddoel waar deze organisaties het veld naar toe proberen te bewegen. Het maakt het mogelijk om vanuit een gedragen positie in het veld bij te dragen aan het uitzetten van richtlijnen ten aanzien van bijvoorbeeld FAIR data principes of het scherper toezien op rechtmatig gebruik van gezondheidsdata. Vanuit het ministerie van VWS moet het initiatief worden genomen om een klankbord groep en/of een directe lijn te openen om het gesprek over kansen en zorgen vanuit het veld bespreekbaar te maken. Hiernaast kan dat de overheid ondersteuning kan bieden bij verschillende vraagstukken waar deze (koepel-)organisaties mee worstelen. Een voorbeeld hiervan is het aanbieden van inhoudelijke ondersteuning door het faciliteren van werksessies met experts en het delen van kennis vanuit andere organisaties die bezig zijn met databeschikbaarheid. Daarnaast kan het ministerie organisatorische ondersteuning bieden door het adviseren van deze partijen over hun lange termijn doelen en kunnen werknemers vanuit het ministerie gedetacheerd worden bij deze constructen om de slagkracht en schaalbaarheid te vergroten. Als laatste kan er structurele financiering geboden worden die zorgt voor een stabiele en gezonde basis waarop deze organisaties lange termijn plannen vorm kunnen geven en vertrouwen krijgen om deze ook uit te voeren. Deze financiering zal gekoppeld moeten zijn aan de lange termijn doelen van deze constructies én het ministerie. Het duidelijk formuleren van het doel van deze financiering tezamen met de verwachten uitkomsten draagt bij aan eenduidige toetsing van de behaalde resultaten.

5.2 VERBREDEN DATABESCHIKBAARHEID

De constructen die zijn meegenomen in deze inventarisatie beslaan slechts een klein deel van de thema's binnen het gezondheidsdomein. Voor succesvol gebruik van AI is het van belang om een kritieke massa (hoeveelheid) van data toegankelijk te maken. Als voorbeeld zijn in deze inventarisatie relatief veel constructies meegenomen die zich richten op data over kanker. Voor analyses specifiek over kanker is dit geen probleem, maar bij analyses die kanker willen vergelijken of koppelen aan andere ziektebeelden wordt het een stuk ingewikkelder als deze data gelimiteerd voorhanden is. Vandaar dat het van belang is om de databeschikbaarheid te verbreden naar meer onderwerpen en thema's. Op dit moment is er nog genoeg ruimte binnen de huidige wettelijke kaders om meer data beschikbaar te maken van andere specifieke aandoeningen of thema's. Het stimuleren van nieuwe- of verbreding van bestaande dataconstructies is dan ook een belangrijke stap naar een grotere databeschikbaarheid voor gezondheid en zorg.

Een centraal aanspreekpunt vanuit de koepelorganisaties in samenwerking met de overheid kan bijdragen aan de verbreding van databeschikbaarheid. Geïnitieerd vanuit de overheid, maar gedragen door het veld biedt een centraal aanspreekpunt een plek waar alle vragen vanuit het veld samen komen en partijen met elkaar worden verbonden. Dit aanspreekpunt is meer dan een loket voor informatie. Het vergt een proactieve benadering vanuit de vertegenwoordigers van dit aanspreekpunt naar het veld om partijen sturing te bieden, te ondersteunen bij regelgeving en het opzoeken van de samenwerking met andere partijen. Het biedt een landingsplek voor nieuwe initiatieven om zich kenbaar te maken aan het veld en om informatie op te halen vanuit ervaringsdeskundigen. Daarnaast biedt dit de mogelijkheid voor constructies om zich te verenigen in werkgroepen rondom verschillende vraagstukken zoals regelgeving, harmonisatie en vertrouwen (zie sectie 4.5). Dit maakt het mogelijk om met gelijkgestemden te spreken over deze thema's, afstemming te zoeken en om met één stem te spreken over deze onderwerpen met partijen buiten het veld. Andersom geeft een centraal aanspreekpunt de mogelijkheid aan onderzoekers, zorginstellingen en commerciële partijen om informatie op te halen over verschillende databeschikbaarheidsinitiatieven. Het wordt dan gemakkelijker om vraag en aanbod van gezondheidsdata op elkaar aan te sluiten waardoor nieuwe en bestaande initiatieven beter kunnen worden getoetst op hun waarde en relevantie binnen het domein gezondheid en zorg.

5.3 VERBINDING MET DE INDUSTRIE

Op dit moment is er slechts beperkte samenwerking tussen dataconstructies en commerciële partijen om data te gebruiken voor AI. In de praktijk zijn veel dataconstructies terughoudend in het delen van data met de industrie vanwege gevreesde wisselende belangen en blijft de databeschikbaarheid beperkt tot wetenschappelijk onderzoek en zorginstellingen. Dit leidt ertoe dat de waarde van data uit de verschillende dataconstructies in het veld deels onbenut blijft doordat zij een nuttige partner mislopen. De industrie is een partner die ervaring heeft met de ontwikkeling van bruikbare AI-toepassingen en ook weet hoe dit naar de praktijk te moeten brengen. Het gaat hier om partijen die kennis kunnen meenemen vanuit andere sectoren en projecten naar het gezondheid- en zorgdomein. De specialisatie van deze partijen op AI zorgt ervoor dat ontwikkeling, trainen, valideren en toepassen sneller kunnen worden doorlopen. In tegenstelling tot zorginstellingen, wetenschappelijk onderzoek en dataconstructies waar gebruik en borging van AI nu niet een hoofddoel is. Samenwerking kan leiden tot informatie-uitwisseling tussen partijen zodat de kennis die aanwezig is in de industrie ook richting dataconstructies kan vloeien. Echter, doordat commerciële partijen op dit moment slechts beperkte toegang hebben tot de waardevolle data die constructies beheren kunnen zij hierdoor niet de potentie van AI voor gezondheid en zorg realiseren. De inventarisatie van AI-toepassingen in gezondheid en zorg in Nederland door KPMG onderschrijft het belang van samenwerking met bedrijven om tot gebruik van AI in gezondheid en zorg te komen³⁵.

Om de verbinding tussen de industrie en de data constructies te stimuleren is het allereerst van belang om het vertrouwen te scheppen dat de belangen van commerciële partijen in lijn liggen met die van de dataconstructies. Dit vertrouwen moet over twee verschillende assen worden gecreëerd. Als start zullen er garanties moeten komen vanuit de industrie om aan te tonen dat zij de datasoevereiniteit en veiligheid kunnen borgen. Het opstellen van toetsingscriteria waaraan commerciële partijen moet voldoen om in aanmerking te komen voor een partnership met een dataconstructie (o.a. security, privacy, ethiek, belangen) zal duidelijkheid scheppen over de minimale vereisten, maar ook bijdragen aan meer samenwerking doordat commerciële partijen kunnen aantonen dat zij voldoen deze eisen. Er is hier een rol om vanuit het centrale aanspreekpunt dit toetsingskader op te stellen waaraan deze partijen moeten voldoen. Ten tweede, zal het veld geïnformeerd moeten worden over de kansen die commerciële partijen bieden voor de toekomst van gezondheid in Nederland. Via informatiebijeenkomsten en werksessies moet deze kennis zich wijder verspreiden en kunnen partijen actief aan elkaar gekoppeld worden. Vanuit het centrale aanspreekpunt moeten deze sessie worden geïnitieerd, in de tussentijd kan hier vanuit het ministerie van VWS al mee worden gestart. Op deze manier raken partijen uit de industrie, die ervaring hebben met AI toepassingen, in contact met de partijen in het veld die de benodigde data voor AI beschikbaar maken. Door verbeterde samenwerking kan kennis over AI en de uitdagingen van databeschikbaarheid tussen deze partijen gedeeld worden. Dit contact is een tweede belangrijke stap om vertrouwen te creëren om gezamenlijk te werken aan AI voor gezondheid en zorg.

5.4 VERTROUWEN UIT DE SAMENLEVING

Veel burgers en patiënten lijken bereid om te handelen volgens het principe van datasolidariteit - het delen van jouw gegevens ter voorkoming of genezing van een ander. Volgens een onderzoek van de Patiënten Federatie Nederland geeft 83% van patiënten aan positief te zijn over gebruik van hun medische gegevens voor wetenschappelijk onderzoek. Echter, uit hetzelfde onderzoek blijkt dat als patiëntgegevens gebruikt worden voor ander soort onderzoek zoals het maken van medicijnen, hulpmiddelen of apps door commerciële organisaties nog maar 36% van de patiënten altijd bereid zijn gegevens te delen. 53% van de patiënten geeft aan dat datadeling met commerciële partijen afhangt van de organisatie en het doel van de organisatie. Veel deelnemers geven geen toestemming voor

³⁵ KPMG (2020), "Inventarisatie AI-toepassingen in gezondheid en zorg in Nederland"
<https://www.rijksoverheid.nl/documenten/rapporten/2020/10/05/inventarisatie-ai-toepassingen-in-gezondheid-en-zorg-in-nederland>

puur commerciële doeleinden of voor doeleinden waar het belang van de patiënt niet op nummer één staat³⁶. Waar dataconstructies minder vertrouwen hebben in het samenwerken met commerciële partijen, hebben burgers en patiënten dat ook. Dit limiteert de dataverzameling aan de bron bij patiënten en burgers en zorgt ervoor dat veel gegevens enkel voor wetenschappelijk onderzoek gebruikt kunnen worden. Het vergroten van het vertrouwen van de burgers en patiënten in het zorgvuldig en veilig gebruiken van medische gegevens door een brede groep van partijen draagt bij aan de databeschikbaarheid.

Het aantonen aan burgers en patiënten dat er voldoende commerciële partijen zijn die ook het belang van de patiënt op nummer één hebben staan gaat hierbij helpen. Zo kan bijvoorbeeld een fabrikant van zorgapparatuur met behulp van meer medische gegevens zijn software beter kalibreren, wat resulteert in betere apparatuur en daarmee betere zorgverlening aan de patiënt. Hierbij gaan commerciële belangen en belangen voor de patiënt hand in hand. Het beter informeren van de patiënt en burger met behulp van praktijkvoorbeelden gaat helpen bij het vergroten van de bereidheid van patiënten om data te delen met commerciële partijen. Het veld mag namelijk niet verwachten van patiënten dat zij hun medische gegevens delen voor een veelvoud aan (wetenschappelijke) onderzoeken, (commerciële) toepassingen en rapporten zonder dat patiënten zelf inzicht hebben in waar hun medische gegevens voor worden gebruikt. Hierbij is een rol weggelegd voor dataconstructies, zorginstellingen en commerciële partijen om burgers en patiënten mee te nemen als doelgroep waaraan de informatie wordt verstrekt. Dit kan op verschillende manieren, zoals de data beschikbaar maken via publiek toegankelijke portalen, het bieden van praktijkvoorbeelden tijdens consultatie met een zorgverlener of via publieke campagnes. Daarnaast moeten praktijkvoorbeelden voor burgers en patiënten naar de voorgrond worden geschoven op websites van dataconstructies die op dit moment in veel gevallen nog altijd zijn gericht op vakgenoten en zorgverleners.

³⁶ Patiënten Federatie Nederland (2019), <https://www.patiëntenfederatie.nl/actueel/nieuws/delen-medische-data-voor-meeste-patiënten-geen-punt>

6. OVERZICHT INITIATIEVEN

6.1 PALGA



Oprichtingsjaar: **1971** / Werknemers: **+ - 10 FTE** / Type data: **Pathologie data**

De Stichting PALGA is een initiatief van pathologen, opgericht in 1971. De missie van PALGA is om alle pathologie informatie te organiseren en bruikbaar te maken voor patiëntenzorg en wetenschap. PALGA beheert de databank met alle Nederlandse pathologie-uitslagen (landelijk dekkend en compleet vanaf 1991) ten behoeve van evaluatie en monitoring van bevolkingsonderzoeken, wetenschappelijk onderzoek, meldingen aan de kankerregistratie, kwaliteitszorg en voor ondersteuning van de directe patiëntzorg. Alle pathologen leggen hun verslagen vast in de PALGA-database, zowel narratief als gestructureerd op basis van protocollen (synoptic reporting). De databank bevatte op 31 december 2019 ruim 76 miljoen pathologie verslagen, waarvan 1,6 miljoen gestructureerde verslagen, en het aantal pathologie verslagen groeit elk jaar met 2,5 miljoen.

Website: www.palga.nl

6.2 HARTWIG MEDICAL FOUNDATION



Oprichtingsjaar: **2015** / Werknemers: **20 FTE** / Type data: **genetische en klinische data van kankerpatiënten met uitgezaaide ziekte**

De stichting Hartwig Medical Foundation is als sequencing- en databasingfaciliteit gestart met filantropie en ontvangt ook projectsubsidies van KWF Kankerbestrijding en ZonMw. De missie is om elke patiënt met uitgezaaide kanker een behandeling op maat te geven, gebaseerd op het DNA van de tumor. Daartoe voert de foundation twee hoofdactiviteiten uit. Allereerst om de juiste behandeling te kunnen geven is een uitgebreide DNA-test ontwikkeld om alle kenmerken van de tumor in kaart te brengen (voor de patiënt van vandaag). Als tweede activiteit faciliteert de stichting onderzoek door derden met het ter beschikking stellen van data uit de databank (voor de patiënt van morgen). Deze data betreft klinische en genetische data van patiënten, die in het kader van klinische studies zijn verkregen via een geïnformeerd consent. Hartwig Medical Foundation heeft de genetische data gegenereerd door het sequensen van lichaamsmateriaal uit biopten en bloedafnames gecollecteerd uit 44 ziekenhuizen. Momenteel bestaat de dataset uit ruim 5000 patiënten en neemt de dataset toe met 200 patiënten per maand. Dit is de grootste data set met whole genome data van patiënten met kanker in de wereld. De data wordt al in meer dan 150 onderzoeken door gerenommeerde kanker centra gebruikt.

Website: www.hartwigmedicalfoundation.nl

6.3 VEKTIS



Oprichtingsjaar: **2006** / Werknemers: **110 FTE** / Type data: **Zorgdeclaratiegegevens**

Vektis C.V. is opgericht in 2006 door en voor de zorgverzekeraars, maar leveren ook informatieproducten voor de overheid, zbo's en zorgaanbieders. Ook ondersteunt Vektis wetenschappelijk onderzoek. Op dit moment bewaart Vektis data over alle zorgdeclaraties in Nederland die wordt aangeleverd door zorgverzekeraars, Wlz-uitvoerders en gegevens van de zorgaanbieders zoals: wachttijden binnen de GG en de onderhanden werkposities van medisch specialistische zorg (MSZ). Daarnaast beheert Vektis het AGB register, wat door Vektis zelf is ontwikkeld. Hierdoor bestaat de database van Vektis uit zorgdeclaraties van alle Nederlanders. De informatie die wordt geleverd en verzameld wordt gebruikt om analyses uit te voeren op de betaalbaarheid en toegankelijkheid van de zorg. Het doel van Vektis is om professionele partijen in de zorg te ondersteunen met op data gebaseerde, feitelijke informatie om de zorg toegankelijk, betaalbaar en van goede kwaliteit te houden. Vektis werkt samen met non-profit organisaties en de data vanuit Vektis draagt bij aan wetenschappelijk onderzoek.

Website: www.vektis.nl

6.4 STATLINE MONITOR LANGDURIGE ZORG



Oprichtingsjaar: **2013** / Werknemers: **3 FTE** / Type data: **Langdurige zorg**

Monitor Langdurige Zorg (MLZ) is een project binnen het team Gezondheid en Zorg van het Centraal Bureau voor de Statistiek (CBS). In opdracht van het Ministerie van VWS haalt CBS datasets binnen over langdurige zorg. Langdurige zorg bestaat uit de verpleging en verzorging van ouderen en chronisch zieken, de zorg voor gehandicapten en de langdurige geestelijke gezondheidszorg. Het CBS combineert deze gegevens met andere informatie beschikbaar binnen het CBS en hierdoor is het mogelijk om nieuwe inzichten te kunnen delen. Data wordt aangeleverd door onder andere de volgende partijen: CIZ, Sociale verzekeringsbank, NZA en CAK. Op dit moment beschrijft de MLZ: de gehele populatie ontvangers van Wlz-gefinancierde zorg, de gehele populatie ontvangers van Zvw-gefinancierde wijkverpleging, zorg voor zintuiglijk gehandicapten en eerstelijnsverblijf en de gehele populatie ontvangers van eigen bijdrage plichtige Wmo-maatwerkarrangementen. Er worden per jaartientallen aanvullende onderzoeksvragen beantwoord op basis van informatie vanuit de MLZ.

Website: <https://www.monitorlangdurigezorg.nl/>

6.5 HEALTH RI



Oprichtingsjaar: **2015** / Werknemers: **8,5 FTE** / Type data: **Gezondheidsdata in brede zin**

Stichting Health RI ondersteund projecten en organisaties die bijdragen aan een nationale gezondheidsdata infrastructuur. De missie van Health RI is om een geïntegreerde onderzoek infrastructuur voor gezondheidsgegevens te bouwen die toegankelijk is voor onderzoekers, burgers en zorgverleners. Het concrete doel van Health-RI is om te functioneren als een one-stop-shop voor onderzoekers in het gepersonaliseerde gezondheidsdomein en daarbuiten. Health-RI zal biobanken, medische beeldvorming en gegevensverzamelingen met elkaar verbinden, waardoor datamining en koppeling mogelijk wordt, en zal toegang bieden tot IT-oplossingen. Naast het delen en verbinden van gegevens, zal Health-RI onderzoeksprojecten ondersteunen met tools voor ethische en juridische diensten, methodologieën, datamanagementtraining, protocollen en best practices. Momenteel zijn er 13 projecten en initiatieven binnen Health RI die bijdragen aan de databeschikbaarheid van gezondheidsdata.

Website: www.healthri.nl

6.6 STICHTING ZORGLADDER



Stichting Zorgladder

Oprichtingsjaar: **2015** / Werknemers: **3 FTE** / Type data: **Uitkomsten van zorgbehandelingen**

Stichting Zorgladder (SZL) is een dataplatform, gestart in 2015 onder de vleugels van ZKN. Sinds 2020 is Zorgladder een stichting. Zorgladder heeft als doel om Uitkomst Gerichte Zorg³⁷ te ondersteunen. Daarnaast heeft Zorgladder als doel om als onafhankelijke stichting en neutrale partij draagvlak te creëren om uitkomstdata landelijk bij elkaar te brengen en toegankelijk te maken. Zorgladder maakt deze uitkomsten zichtbaar doormiddel van dashboards met stuur- en spiegelinformatie voor in de behandel- en bestuurskamer. Zorgladder geeft aan dat het huidige aantal observaties in de dataset onbekend is.

Website: www.zorgladder.nl

6.7 PROTRAIT

Oprichtingsjaar: **2017** / Werknemers: **+/- 10 FTE** / Type data: **alle kankersoorten waarbij bestraling wordt toegepast en waarvoor protontherapie effectief kan zijn**

ProTrait is opgericht in 2017 en gesubsidieerd door KWF Kankerbestrijding. Het doel van ProTrait is dan om een landelijke registratie voor protontherapie op te zetten. De meeste registraties die nu worden gebruikt worden met de hand verzameld. Het doel van ProTrait is om dit te veranderen en deze data FAIR te maken en automatisch te verzamelen, beschikbaar stellen en te analyseren via de Personal Health Train. Eén van de gewenste uitkomsten voor ProTrait is dat voor zorginstellingen dit de registratielast verminderd. Het verwachte resultaat hiervan is dat het voor artsen mogelijk wordt om met betere data onderzoek te kunnen doen wat de kwaliteit van zorg ten goede komt. ProTrait geeft aan dat het huidige aantal observaties in de dataset nog onbekend is. De patiëntenregistratie verplicht, waardoor de dataset blijft toenemen.

Website: (nog) niet aanwezig

³⁷ <https://www.uitkomstgerichte Zorg.nl/>

6.8 PERINED-DATABANK

Oprichtingsjaar: **1971** / Werknemers: **12 FTE** / Type data: **Prenatale zorg**

Stichting Perined beheert een registratie waarin gegevens staan van zwangere vrouwen en pasgeborenen (t/m dag 28). De Perinatale Registratie bevat data van vier beroepsgroepen die zich bezighouden met geboortezorg: verloskundigen, huisartsen, gynaecologen en kinderartsen. Deze registratie wordt door Perined beheerd namens vier beroepsverenigingen: de KNOV, de LHV, de NVOG en de NVK. Indien een zwangere vrouw en/of pasgeborene in zorg is bij meerdere type zorgverleners, dan wordt door elke type zorgverlener gegevens geregistreerd in een elektronisch dossier van een softwareleverancier, enkel die gegevens van zorg waarbij dat type zorgverlener verantwoordelijk was. Perined heeft tot doel om de kwaliteit van de perinatale zorgverlening in Nederland te vergroten door meten en bespreken. Perined heeft met de zorgverleners en haar leveranciers een (sub)verwerkersovereenkomst afgesloten. De afgelopen paar jaar zijn er 160.000 geboortes geregistreerd, welke onderdeel zijn van de dataset van Perined.

Website: www.perined.nl

6.9 STIZON



Oprichtingsjaar: **2012** / Werknemers: **6 FTE** / Type data: **Uitkomsten zorg van geneesmiddelen**

De Stichting Informatievoorziening voor Zorg en Onderzoek (STIZON) komt voort uit het PHARMO instituut. Dit instituut doet onderzoek naar de werking en veiligheid van geneesmiddelen in de praktijk. In 2012 is in overleg met privacy deskundigen, besloten de gegevens die in de loop der jaren zijn verzameld in beheer te geven van een onafhankelijke stichting van zorgverleners en patiënten. STIZON volgt grote groepen patiënten die geneesmiddelen gebruiken langere tijd om te leren wat er misgaat tijdens de behandeling, waarom dit misgaat en hoe dit te voorkomen valt in de toekomst. Het doel van STIZON is om op deze manier bij te dragen aan een betere kwaliteit zorg. STIZON fungeert als een trusted third party (TTP). Dat wil zeggen dat STIZON alleen de data beheert. De analyses op deze data wordt gedaan door INSZO, hierdoor wordt het beheer en de analyses gescheiden. Momenteel bestaat de dataset uit meer dan 11 miljoen observaties.

Website: www.stizon.nl

6.10 LIFELINES BIOBANK



Oprichtingsjaar: **2006** / Werknemers: **45 FTE** / Type data: **geen specifiek ziektebeeld**

LifeLines B.V. verzamelt data ten behoeve van gezondheidsvraagstukken, specifiek gericht op healthy aging. Deze data wordt gekoppeld met externe partijen en beschikbaar gesteld voor wetenschap en beleid. Door terugkoppeling van gezondheidsdata uit de studie zijn afwijkingen vroegtijdig aan het licht gekomen waardoor vroegtijdige behandeling mogelijk was. Voor burgers uit bijvoorbeeld provincie Groningen heeft de publicatie op basis van gegevens uit Lifelines geleid tot een inzicht in de directe relatie tussen aardbevingsschade en gezondheid. Onderzoekers die gebruik willen maken van de data binnen Lifelines Biobank werken zoveel mogelijk in een afgeschermd omgeving. Werkt de onderzoeker buiten deze omgeving, dan wordt er door de onderzoeker een code of conduct getekend. De dataset bestaat momenteel uit 167.000 deelnemers uit Noord-Nederland.

Website: www.lifelines.nl

6.11 BBMRI



Oprichtingsjaar: **2010** / Werknemers: **50 FTE** / Type data: **Biobank samples, afbeelding en data**

Stichting BBMRI is opgericht in 2010 en heeft als doel om het gebruik van biosamples, afbeeldingen en data voor gezondheidsonderzoek voor preventie diagnoses en behandeling van ziektes te maximaliseren. Daarnaast willen BBMRI de toegang tot samples, afbeeldingen en data zo eenvoudig mogelijk maken voor zowel provider als gebruiker. De databeschikbaarheid vanuit BBMRI leidt tot nieuwe kennis waarmee de zorg verbeterd kan worden. Daarnaast stimuleren BBMRI het belang van gelijke en gedeelde datakwaliteit doormiddel van het adopteren van de FAIR principes. Ook de tools voor dataverzameling en opslag, alsmede de benodigde tools voor het uitvoeren van onderzoek, maakt BBMRI beschikbaar. Data binnen het BBMRI wordt aangeleverd door onder andere UMC's, Palga en universiteiten.

Website: www.bbmri.nl

6.12 CANCERDATA.ORG

Oprichtingsjaar: **2010** / Werknemers: **1 FTE** / Type data: **data met betrekking tot karakteristieken bij kankerpatiënten, tumoren en behandelingen**

CancerData.org is onderdeel van de Medical Informatics and Knowledge Engineering team (MIKE) van de Stichting Maastricht in Maastricht. CancerData.org is in 2010 opgericht en heeft als doel om research findings, anonieme Life Science data, intekenatlassen, protocollen en ontologieën publiekelijk te stellen ten behoeve van vervolgonderzoek (voor onderzoekers) en standaardisatie van (radiotherapie) behandelingen voor professionals. De site biedt een centraal, online archief voor de opslag van klinische protocollen, publicaties en onderzoeksdatasets. Primair voor zorgprofessionals betekent dit: consensus van protocollen, intekenatlassen en additionele data bij publicaties. Door volgens gelijke protocollen te werken, zijn data compatible/vergelijkbaar/semantic-interoperable. Dit levert betere data en dus (onderzoeks)resultaten en uiteindelijk betere behandelingen op. Voor patiënten betekent dit: door delen het van datasets wordt via onderzoek gewerkt aan implementatie van predicatieve modellen die voor beslismiddelen worden gebruikt. Momenteel zijn er 34 datasets beschikbaar binnen CancerData.org.

Website: www.cancerdata.org

6.13 AMSTERDAM MEDICAL DATA SCIENCE (AMSTERDAM UMCDB)



Oprichtingsjaar: **2018** / Werknemers: **4 FTE** / Type data: **Intensive care opnames**

Amsterdam Medical Data Science is opgericht in 2018 en heeft als hoofddoel om de zorg te verbinden met data science. De samenwerking met het Nederlandse Vennootschap Intensive Care (NVIC), haar Research Network (RCCnet) en het Amsterdamse UMC heeft geleid tot de productie van Amsterdam UMCdb. Dit initiatief heeft daarnaast als doel om zorgdata met betrekking tot intensive care patiënten te delen onder heel Europa. Dit is de eerste vrij toegankelijke database voor intensive care vanuit de Europese Unie. Als voorwaarde voor het verstrekken van informatie dient er altijd een intensivist deel te nemen aan het onderzoek. De dataset bevat momenteel geïdentificeerde gezondheidsgegevens met betrekking tot 23.000 opnames van Europese intensive care eenheden, waaronder demografische gegevens, vitale functies, laboratoriumonderzoeken en medicijnen.

Website: <https://www.amsterdammedicaldatascience.nl/>

6.14 DUTCH HOSPITAL DATA



Oprichtingsjaar: **2018** / Werknemers: **65 FTE** / Type data: **Ziekenhuisopnames**

Stichting Dutch Hospital Data (DHD) is opgericht vanuit de algemene ziekenhuizen om regie te krijgen op data verzameling van landelijke medische registratie. De belangrijkste is hierbij de LBZ (Landelijke basisregistratie ziekenhuiszorg) dataset van alle opnames in ziekenhuizen (opnames met overnachting en dagopnames). De tweede vraag waaraan wordt voldaan is op basis van deze informatie informatieproducten te maken (bijv. sterftecijfers berekenen). Alle NFZ en NFU leden krijgen benchmarks over deze informatie zodat zij hun bedrijfsvoering kunnen verbeteren. Deze informatieverzameling is voor twee doeleinden. Enerzijds een verantwoording naar de overheid. Anderzijds tot het komen tot een verbetering van de zorg. Daarnaast kan DHD ook advies verstrekken over standaardisatie naar ziekenhuizen, hierbij wordt ondersteuning geboden aan werknemers in het ziekenhuis hoe deze data het best kan worden vastgelegd. De kwaliteit van de data wordt geborgd doordat DHD de data aanlevert voor kwaliteitsregistratie als één partij. Momenteel bestaat de dataset uit alle opnames van patiënten in Nederlandse ziekenhuizen (\pm 75 partijen).

Website: <https://www.dhd.nl/>

6.15 E/MTIC HEALTH DATA PORTAL



Oprichtingsjaar: **2020 (portaal is nog niet live)** / Werknemers: **+ - 10/15 FTE** / Type data: **geen specifiek ziektebeeld**

Health Data Portal wil zorgen voor een platform voor gegevensuitwisseling binnen e/MTIC dat een efficiënte uitwisseling van gezondheidsgegevens mogelijk maakt tussen de onderzoekers, waardoor een infrastructuur ontstaat die kan leiden tot een sneller levenswetenschappen & gezondheidsinnovatie. Het doel is om grote hoeveelheden data uit verschillende bronnen veilig en betrouwbaar en privacy-respecterend op te slaan, te delen en te onderzoeken. Hiervoor wordt een veilige infrastructuur gecreëerd waarin grote datasets met medische informatie (verzameld uit verschillende bronnen en in verschillende formaten) kunnen worden geanalyseerd, onderhouden, uitgebreid en gedeeld tussen de leden van het platform. Het portaal heeft niet als doel om de data centraal op te slaan, maar zal decentrale informatie vindbaar en beschikbaar maken. Op dit moment is de Health Data Portal goedgekeurd om ontwikkeld te worden.

Website: <https://www.tue.nl/en/research/research-groups/eindhoven-medtech-innovation-center/>

6.16 IKNL



Oprichtingsjaar: **1989** / Werknemers: **550 FTE** / Type data: **data met betrekking tot onder andere tumorkarakteristieken, patiëntkarakteristieken en behandelingen van kankerpatiënten**

De Nederlandse Kankerregistratie is 30 jaar geleden opgericht en vormt een bron van betrouwbare en relevante gegevens van kankerpatiënten uit heel Nederland. IKNL ontvangt jaarlijks een instellingssubsidie van het Ministerie van VWS en haalt daarnaast inkomsten uit onderzoeken en projecten. De NKR bevat informatie over diagnostiek, diagnose, tumorkarakteristieken en initiële behandeling, ongeacht de behandellocatie. Het gaat hier voornamelijk om gegevens in de vorm van codes, numerieke waarden en booleaans. Deze gegevens worden geregistreerd in Nederlandse ziekenhuizen en door datamanagers vertaald en gevalideerd vanuit de EPDs naar uniforme en gestandaardiseerde datasets. De gegevens in de NKR worden gebruikt voor wetenschappelijk onderzoek en statistiek naar kanker en de behandeling ervan. Dit wordt vaak gedaan in samenwerking met collega onderzoeksinstellingen, wat leidt tot verbetering van de behandeling van de ziekte en wat uiteindelijk dus aansluit bij de missie van IKNL. Op het moment worden alle data centraal opgeslagen, maar IKNL experimenteert met modellen waarbij de analyse naar de data gaat (Personal Health Train). Momenteel zitten er zo ongeveer 2,5 miljoen casussen in de dataset met 200 datapunten per casus.

Website: <https://www.iknl.nl/nkr>

6.17 LIME



Oprichtingsjaar: **2016** / Werknemers: **5 FTE** / Type data: **geen specifiek ziektebeeld**

LIME (Limburg Meet) is opgericht in 2016 en bevat data van de Limburgse bevolking. LIME is opgezet om gezondheidsdata slimmer te meten en deze data efficiënt te analyseren. Het uiteindelijke doel is om de gezondheid van inwoners in Limburg te verbeteren. LIME is een team van experts dat indieners van vragen in contact kan brengen met relevante partners uit het netwerk van LIME. LIME wil alle aspecten van de zorg maar ook andere sociaal economische aspecten betrekken om een integraal beeld te creëren. De data wordt dan ook vanuit verschillende instanties verzameld, zoals onder andere het CBS, scholen, gemeentes, wijkverpleeghuizen, huisartsen, ziekenhuizen, UMC's en psychologen. De dataset van LIME bestaat uit 10,000 observaties afkomstig uit Maastricht, welke wordt gekoppeld aan de zeventien miljoen observaties afkomstig uit CBS. Onderzoekers die gebruikt willen maken van de data in het netwerk van LIME gebruiken hiervoor het Health Data innovatie lab van LIME.

Website: www.limeconnect.nl

6.18 NIVEL



Oprichtingsjaar: **1985** / Werknemers: **150 medewerkers** / Type data: **Zorgregistraties eerste lijn, beroepenregistraties en verschillende panels**

Het Nivel levert de kennis om de gezondheidszorg in Nederland beter te maken met hoogwaardig, betrouwbaar en onafhankelijk wetenschappelijk onderzoek naar thema's met een groot maatschappelijk belang. Onderzoek van het Nivel omvat het perspectief van patiënten, cliënten en burgers; het perspectief van zorgprofessionals en organisaties ; en het perspectief van het gezondheidszorgsysteem. Het Nivel onderhoudt verschillende databases: Nivel Zorgregistraties Eerste Lijn, beroepen in de gezondheidszorg, de database communicatie in de gezondheidszorg, de databank Communicatie in de Gezondheidszorg, het panel verpleging en verzorging, het consumentenpanel gezondheidszorg, nationaal panel chronisch zieken, het panel samen leven, het panel mensen met een longziekte, en burgerplatforms. Samen vormen deze een rijke bron van gegevens over zorg en gezondheid. De omvangrijkste database is Nivel Zorgregistraties Eerste Lijn, waarin gegevens zijn verwerkt uit elektronische patiëntendossiers van huisartsen en ander eerstelijnszorgverleners. Zo omvat deze database longitudinale huisartsengegevens uit ongeveer 10% van de Nederlandse bevolking (1,7 miljoen mensen), met gegevens over gezondheid, zorggebruik, verwijzingen naar de medisch specialist, medicatie. Een ideale basis voor *artificial intelligence* en *machine learning* toepassingen. Zo werden Nivel data gebruikt voor voorspellingsmodellen voor suïcide, en beslissingsondersteuning bij het voorschrijven van geneesmiddelen in de huisartsenpraktijk. Maar ook om besmettingen met Covid-19 door de tijd te volgen, of patronen van zorggebruik te onderzoeken. Ook derden kunnen gegevens aanvragen voor nadere analyses.

Website: www.nivel.nl

6.19 NVVR-NCIF BEELDBANK I.O

Oprichtingsjaar: **in oprichting** / Werknemers: **n.v.t.** / Type data: **radiologie beelden omtrent COVID-19**

De Nederlandse Vereniging voor Radiologie is de beroepsvereniging van radiologen in Nederland en is het Nederland COVID-19 Initiatief (NCIF) gestart om landelijk radiologische beelden, verslagen en patiëntgegevens te verzamelen, spiegelen en onderzoeken. Het NCIF is voor de NVvR een proeve van een landelijke BeeldBank Radiologie en is opgericht vanuit een vraag die is ontstaan vanuit Nederlandse ziekenhuizen en vanuit commerciële partijen naar beschikbaarheid van radiologiebeelden. Het initiatief streeft naar een veilige en betrouwbare landelijk dekkende data-voorziening om met toestemming van de patiënt kennis te delen en ook op deze manier onderzoek te faciliteren. NCIF heeft als doel om een database op te zetten met in samenwerking met meerdere specialismen, waardoor het mogelijk is diverse vraagstukken ten behoeve van COVID-19 aan te pakken. NCIF zoekt via de radiologie afdelingen in academische centra toenadering tot andere onderzoekers, waarbij expertise op het gebied van imaging en artificiële intelligentie (AI) beschikbaar is. Momenteel zijn er binnen verschillende ziekenhuizen initiatieven om patiënttoestemming rondom COVID-19 vast te leggen om deze gegevens te kunnen gebruiken voor wetenschappelijk onderzoek.

Website: www.radiologen.nl/ncif-beeldbank-radiologie-oprichting

6.20 HIV MONITORING



Opgericht: **2002** / Werknemers: **+ - 45 FTE** / Type data: **alle biomedische data van HIV positieve patiënten**

Stichting HIV Monitoring (SHM) is opgericht in 2002 en verzamelt data uit alle HIV behandelcentra in Nederland en incidenteel uit andere kwaliteitsregistraties in Nederland. De stichting wordt gefinancierd door het Ministerie van VWS. SHM heeft meerdere doelen. Een eerste doelstelling waar de stichting zich op focust is het verstrekken van informatie aan alle belanghebbende organisaties, waaronder zorgverleners, de overheid, onderzoekers en mensen die leven met hiv over landelijke en regionale trends in alle aspecten van de hiv-infectie. Hiertoe behoort, naast informatie over het beloop van de hiv-epidemie en de uitkomsten van behandeling van hiv-infecties, onder andere ook informatie over comorbiditeiten en co-infecties, inclusief virale hepatitis, in hiv-positieve personen in zorg in Nederland. Verder wil de stichting modellen ontwikkelen die toekomstige trends in de hiv-epidemie en in het klinische beloop van hiv-positieve personen in zorg in Nederland accuraat kunnen voorspellen. Ook het monitoren van en rapporteren over de kwaliteit van de hiv-zorg in Nederland, behoort tot de doelstellingen van SHM. Hiermee wordt bijgedragen aan een nationale standaard voor de kwaliteit van hiv-zorg en aan de formele certificatie van hiv-behandelcentra in Nederland. Tot slot wil de stichting bijdragen aan nationale en internationale wetenschappelijke onderzoekssamenwerkingen en fungeren als nationaal kenniscentrum voor informatie over trends van alle relevante aspecten van de hiv-infectie en het klinische beloop van hiv-positieve personen in zorg in Nederland. Momenteel bestaat de dataset uit 31,070 participanten.

6.21 STICHTING FARMACEUTISCHE KENGETALLEN



Oprichtingsjaar: **1990** / Werknemers: **6,5 FTE** / Type data: **data met betrekking tot geneesmiddelen**

De Stichting Farmaceutische Kengetallen (SFK) is opgezet naast de KNMP (Koninklijke Nederlandse Maatschappij ter bevordering der Farmacie) en houdt zich sinds 1990 bezig met het verzamelen van gedetailleerde gegevens omtrent het geneesmiddelengebruik in Nederland. De stichting bevordert goede geneesmiddelenvoorziening, de wetenschappelijke beoefening van de farmacie en de belangenbehartiging van openbare apothekers in Nederland. De SFK geeft ook inzicht in de omvang en opbouw van de geneesmiddelenvoorziening via de openbare apotheken om apothekers ten behoeven van hun managementtaak in de apotheek van dienst te zijn. De SFK beschikt over een omvangrijke en gedetailleerde gegevensverzameling op het gebied van geneesmiddelen gebruik in Nederland. De SFK is gefinancierd door het Ministerie van VWS en bevat zestien miljoen observaties in de dataset.

Website: www.sfk.nl

6.22 DUTCH INSTITUTE FOR CLINICAL AUDITING)



Oprichtingsjaar: **2010** / Werknemers: **40 FTE** / Type data: **DICA bevat 22 kwaliteitsregistraties uiteenlopend van obesitas tot aan longkanker**

DICA is opgericht in 2010 door artsen en is door de jaren heen een uitgebreidere en bredere organisatie geworden in de gezondheidszorg. DICA kijkt naar specifieke aandoeningen om hier vervolgens kwaliteitsregistraties van te maken, inmiddels telt DICA in totaal 22 kwaliteitsregistraties. Een aantal voorbeelden van deze kwaliteitsregistraties zijn Dutch Acute Stroke Audit, Dutch Audit for Treatment of Obesity en Dutch Breast Implant Registry. Door ziekenhuizen in Nederland worden verschillende definities verzameld over deze aandoeningen, waarbij er tussen de verschillende ziekenhuizen is afgesproken deze definities op dezelfde wijze te documenteren. De ziekenhuizen leveren deze data aan bij MRDM, zodat deze organisatie de data kan pseudonimiseren en door kan sturen naar DICA. DICA verwerkt deze data in een zogenaamd dashboard en laat hier diverse analyses op los, zodat deze analyses daarna weer teruggestuurd worden naar de artsen in de ziekenhuizen. Door deze dashboards op te stellen maakt DICA de kwaliteit van zorg inzichtelijk met betrouwbare vergelijkingen en analyses. Hierbij wordt samengewerkt met dokters, patiënten en zorgverzekeraars. Deze zorgverzekeraars zijn voor DICA dan ook de voornaamste inkomensbron. Naast de zorgverzekeraars ontvangt DICA ook financiering uit diverse onderzoeksprojecten.

Website: www.dica.nl

6.23 NEDERLANDSE HARTREGISTRATIE



Oprichtingsjaar: **2017** / Werknemers: **10 FTE** / Type data: **kwaliteitsregistraties met betrekking tot hartinterventies en hartoperaties**

De stichting Nederlandse Hart Registratie (NHR) is een kwaliteits- en registratie organisatie waar verschillende registraties van alle hartinterventies en hartoperaties in Nederland samenkomen. De doelstelling van de NHR is bijdragen aan het behoud en verdere verbetering van kwaliteit van zorg; door het verzamelen, analyseren en terugkoppelen van relevante gegevens over de behandeling van patiënten met hartaandoeningen. Door middel van deze registratie wordt de hartpatiënt door alle fases van het behandeltraject heen gevolgd, vanaf het moment van stellen van de diagnose tot jaren na de interventie. De NHR wordt gefinancierd door de contributies die zij ontvangen vanuit de deelnemende ziekenhuizen. Momenteel bevat de dataset van NHR twee miljoen observaties uit 16 hartcentra, 43 algemene ziekenhuizen en 14 dottercentra.

Website: <https://nederlandsehartregistratie.nl/>

7. APPENDIX

7.1 REFERENTIES

1. Advies commissie governance van kwaliteitsregistraties (2019), <https://www.rijksoverheid.nl/documenten/rapporten/2019/07/05/advies-commissie-governance-van-kwaliteitsregistraties>
2. Autoriteit Persoonsgegevens (2020), <https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/algemene-informatie-avg/mag-u-persoonsgegevens-verwerken#wat-verstaat-de-avg-onder-bijzondere-persoonsgegevens-6339>
3. CBS, CBS Statline (2020) <https://opendata.cbs.nl/#/CBS/nl/>
4. Europese Commissie (2020). "Witboek over kunstmatige intelligentie - een Europese benadering op basis van excellentie en vertrouwen" https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_nl.pdf
5. Europese Commissie (2020). Europese Commissie (2020). "Een Europese datastrategie" <https://eur-lex.europa.eu/legal-content/NL/TXT/HTML/?uri=CELEX:52020DC0066&from=NL>
6. Kamerbrief over visie op medische technologie (2019), <https://www.rijksoverheid.nl/documenten/kamerstukken/2019/04/26/kamerbrief-over-visie-op-medische-technologie>
7. KPMG, Inventarisatie AI-toepassingen in gezondheid en zorg in Nederland (2020), <https://www.rijksoverheid.nl/documenten/rapporten/2020/10/05/inventarisatie-ai-toepassingen-in-gezondheid-en-zorg-in-nederland>
8. Ministerie van Volksgezondheid, Welzijn en Sport (2018). "Data laten werken voor gezondheid" <https://www.rijksoverheid.nl/documenten/brieven/2018/11/15/data-laten-werken-voor-gezondheid>
9. Moons et al. (2014) "Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist" <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001744>
10. Nederlandse AI Coalitie (NLAIC) (2020). "Verantwoord datadelen voor AI" <https://nlaic.com/wp-content/uploads/2020/03/Verantwoord-datadelen-voor-AI.pdf>
11. Nivel & CBS (2019) "Hergebruik van routine zorgdata voor beleid en wetenschap: hoe het nog beter kan" <https://www.nivel.nl/nl/publicatie/hergebruik-van-routine-zorgdata-voor-beleid-en-wetenschap-hoe-het-nog-beter-kan>
12. NZA, DIS (2020) <https://www.opendisdata.nl/dis/over>
13. Patiënten Federatie Nederland (2019), "Delen van gezondheidsgegevens voor onderzoek" <https://www.patiëntenfederatie.nl/downloads/monitor/153-delen-van-gezondheidsgegevens-voor-onderzoek/file>
14. Personal Health Train (2020) <https://pht.health-ri.nl/nl>
15. RIVM, RIVM Statline (2020) <https://statline.rivm.nl/#/RIVM/nl/>
16. ZonMW, FAIR data and data management (2020) <https://www.zonmw.nl/en/research-and-results/fair-data-and-data-management/>
17. Zorginstituut Nederland (2018). "Eindrapportage praktijktoets FAIR Data & Personal Health Train" <https://www.zorginstituutnederland.nl/publicaties/rapport/2018/10/23/eindrapportage-praktijktoets-personal-health-train>
18. ZorgTTP (2020) <https://www.zorgttp.nl/trusted-third-party-ttp/>

Over Capgemini Invent

Als het digitale innovatie-, advies- en transformatiemark van de Capgemini Group, helpt Capgemini Invent CxO's bij het bedenken en bouwen van wat de toekomst brengt voor hun organisaties. Met meer dan 30 kantoren en 25 creatieve studio's over de hele wereld, combineert het 7.000+ sterke team strategie, technologie, datawetenschap en creative design met diepgaande branchekennis en inzichten om nieuwe digitale oplossingen en de bedrijfsmodellen van de toekomst te ontwikkelen.

Capgemini Invent is een onderdeel van Capgemini, een wereldleider op het gebied van consulting en technologie diensten en digitale transformatie. Capgemini loopt voorop als het gaat om innovatie en ondersteunt klanten in de veranderende wereld van cloud, digitaal en platforms. Voortbouwend op 50 jaar geschiedenis en diepgaande branche specifieke ervaring, stelt Capgemini organisaties in staat hun zakelijke ambities te realiseren via een reeks van diensten van strategisch advies tot IT implementaties op de werkvloer.

Capgemini is een multicultureel bedrijf van meer dan 200.000 teamleden in meer dan 40 landen.

Erik Hoorweg

Vice President Public Sector
Capgemini Invent NL
T: +31 615030869
E: erik.hoorweg@capgemini.com

Willem Sträter

Manager Data for Intelligent Operations
Capgemini Invent NL
T: +31 652734288
E: willem.strater@capgemini.com

Voor meer informatie:
www.capgemini.com/invent

